

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:13 : Search time 303.5 seconds
(without alignments)
2.441 Million cell updates/sec

Title: US-09-446-109a-3
Perfect score: 50
Sequence: 1 YSFKDPLXR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDSB/gcgdata/geneseq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	98.0	10	17	AA94470
2	49	98.0	10	17	AA94473
3	48	96.0	10	17	AA94476
4	48	96.0	10	17	AA94468
5	47	94.0	10	17	AA94477
6	44	88.0	10	17	AA94475
7	43	86.0	10	17	AA94474
8	43	86.0	10	17	AA94482
9	42	84.0	10	17	AA94478
10	42	84.0	10	17	AA94479
11	42	84.0	10	18	AAW27250

12	42	84.0	12	18	AAW27254	Molecular adjuvant
13	42	84.0	18	18	AAW27256	Molecular adjuvant
14	42	84.0	19	18	AAW27251	Molecular adjuvant
15	41	82.0	10	17	AA94486	C5a anaphylatoxin
16	41	82.0	10	17	AA94489	C5a anaphylatoxin
17	41	82.0	10	17	AA94490	C5a anaphylatoxin
18	41	82.0	10	17	AA94492	C5a anaphylatoxin
19	41	82.0	10	17	AA94493	C5a anaphylatoxin
20	41	82.0	10	17	AA94494	C5a anaphylatoxin
21	41	82.0	10	17	AA94495	C5a anaphylatoxin
22	41	82.0	10	17	AA94496	C5a analogue C5a(6
23	41	82.0	10	17	AA94497	C5a analogue C5a(6
24	41	82.0	10	17	AA94498	C5a analogue C5a(6
25	41	82.0	10	17	AA94499	C5a analogue C5a(6
26	41	82.0	10	17	AA94500	C5a analogue C5a(6
27	41	82.0	10	17	AA94471	C5a anaphylatoxin
28	41	82.0	10	17	AA94480	C5a anaphylatoxin
29	41	82.0	10	17	AA94481	C5a anaphylatoxin
30	40	80.0	9	17	AA94469	C5a anaphylatoxin
31	40	80.0	10	17	AA94488	C5a anaphylatoxin
32	40	80.0	10	17	AA94465	C5a anaphylatoxin
33	39	78.0	10	17	AA94466	C5a anaphylatoxin
34	38	76.0	10	17	AA94472	C5a anaphylatoxin
35	36	72.0	521	19	AAW47237	Human lymphoid-spe
36	36	72.0	581	19	AAW47238	Human lymphoid-spe
37	34	68.0	10	17	AA94487	C5a anaphylatoxin
38	33	66.0	10	16	AA94464	Control decapeptid
39	33	66.0	10	17	AA94464	Human C5a anaphyla
40	33	66.0	10	17	AA94467	C5a anaphylatoxin
41	33	66.0	517	22	AAE05394	Staphylococcus aur
42	33	66.0	521	22	AAE05392	Staphylococcus aur
43	33	66.0	546	18	AAW20076	Hexose oxidase, an
44	33	66.0	546	21	AA93619	Synthetic hexose o
45	33	66.0	546	22	AA59205	Chondrus crispus he

ALIGNMENTS

RESULT 1
AAR94470
ID AAR94470 standard; peptide; 10 AA.
XX
AC AAR94470;
XX
XX
DT 17-OCT-1996 (first entry)
XX
DE C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 Ala73.
XX
XX C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
KW constrained backbone conformation; proinflammatory response;
KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
KW mediated; increase; cell membrane; vascular; permeability;
KW neutrophil polarisation; neutrophil enzyme release; treatment;
KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
KW cancer; drug delivery; CNS disorder; central nervous system;
KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
KW development; non-steroidal; anti-inflammatory.
XX
OS Homo sapiens.

Key	Location/Qualifiers
Region	4..7
FT	/note= "beta-turn"
XX	
PN	WO9606629-A1.
XX	
PD	07-MAR-1996.
XX	
PF	31-AUG-1995; 95WO-US11126.
XX	
PR	31-AUG-1994; 94US-0299285.
XX	

PA (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX MPI; 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 XX spasmogenesis, platelet aggregation and increases in cell membrane
 XX permeability
 XX Example 2: Page 79; 116pp; English.
 XX The present peptide is an analogue of a C-terminal decapeptide
 XX region from human C5a anaphylatoxin, which has a constrained
 XX backbone conformation (compared to the natural peptide's flexible
 XX C-terminal region) comprising a beta-turn. Analogues with type
 XX (II or V) and type III beta-turns elicit proinflammatory responses
 XX characterised by spasmogenesis, platelet aggregation and neutrophil
 XX non-mediated increases of cell membrane (specifically vascular)
 XX permeability, or neutrophil polarisation, neutrophil enzyme
 XX release and neutrophil mediated increases in cell membrane (esp.
 XX vascular) permeability, respectively.
 XX The analogues can be used as immune adjuvants for the treatment of
 XX immune deficiency disorders, and for augmenting standard immune
 XX therapy for the treatment of cancer, without inflammatory side
 XX effects. They can also be used to facilitate drug delivery by
 XX increasing vascular permeability, e.g. in the treatment of CNS
 XX disorders such as Alzheimer's disease or tumours, and to develop
 XX high affinity C5a receptor antagonists useful as non-steroidal
 XX anti-inflammatory agents.
 XX In a type (II or V) beta-turn specific human foetal artery smooth
 XX muscle contraction assay, and a type III beta-turn specific
 XX neutrophil polarisation assay the peptide had respective EC50
 XX (microm) activities of 122 and 48, compared to 0.018 and 0.0013
 XX for C5a.
 XX SInfluence 10 AA;
 Query Match 98.0%; Score 49; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.0013;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YSFKDMPLXR 10
 Db 1 ysfkdmplar 10
 RESULT 2
 AAR94473 ID AAR94473 standard; peptide: 10 AA.
 AC AAR94473;
 XX 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 Pro71 D-Ala73.
 XX C-terminal; human: C5a anaphylatoxin; analogue: beta-turn;
 XX constrained backbone conformation; proinflammatory response;
 XX spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 XX mediated; increase: cell membrane; vascular; permeability;
 XX neutrophil polarisation; neutrophil enzyme release; treatment;
 XX immune adjuvant; immunodeficiency; augmentation; immune therapy;
 XX cancer; drug delivery; CNS disorder; central nervous system;
 XX Alzheimer's disease; tumour; high affinity; receptor antagonist;
 XX development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 XX Region 4..7

FT Misc-difference 9 /note= "beta-turn"
 FT /note= "D-form residue"
 XX WO9606629-A1.
 XX 07-MAR-1996.
 XX 31-AUG-1995; 95WO-US11126.
 XX 31-AUG-1994; 94US-0299285.
 XX (UYNE-) UNIV NEBRASKA.
 XX (UYOU) UNIV QUEENSLAND.
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX MPI; 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 XX spasmogenesis, platelet aggregation and increases in cell membrane
 XX permeability
 XX Claim 8: Page 81; 116pp; English.
 XX The present peptide is an analogue of a C-terminal decapeptide
 XX region from human C5a anaphylatoxin, which has a constrained
 XX backbone conformation (compared to the natural peptide's flexible
 XX C-terminal region) comprising a beta-turn. Analogues with type
 XX (II or V) and type III beta-turns elicit proinflammatory responses
 XX characterised by spasmogenesis, platelet aggregation and neutrophil
 XX non-mediated increases of cell membrane (specifically vascular)
 XX permeability, or neutrophil polarisation, neutrophil enzyme
 XX release and neutrophil mediated increases in cell membrane (esp.
 XX vascular) permeability, respectively.
 XX The analogues can be used as immune adjuvants for the treatment of
 XX immune deficiency disorders, and for augmenting standard immune
 XX therapy for the treatment of cancer, without inflammatory side
 XX effects. They can also be used to facilitate drug delivery by
 XX increasing vascular permeability, e.g. in the treatment of CNS
 XX disorders such as Alzheimer's disease or tumours, and to develop
 XX high affinity C5a receptor antagonists useful as non-steroidal
 XX anti-inflammatory agents.
 XX In a type (II or V) beta-turn specific human foetal artery smooth
 XX muscle contraction assay, and a type III beta-turn specific
 XX neutrophil polarisation assay the peptide had respective EC50
 XX (microm) activities of 0.35 and 2.1, compared to 0.018 and 0.0013
 XX for C5a.
 XX Sequence 10 AA;
 Query Match 98.0%; Score 49; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.0013;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 YSFKDMPLXR 10
 Db 1 ysfkdmplar 10
 RESULT 3
 AAR94476 ID AAR94476 standard; peptide: 10 AA.
 XX AAR94476;
 XX 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 Pro71 D-Phe73.
 XX C-terminal; human: C5a anaphylatoxin; analogue: beta-turn;
 XX constrained backbone conformation; proinflammatory response;

KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX
 OS Homo sapiens.
 XX
 FH Key "Location/Qualifiers
 FT Region 4..7
 FT Misc-difference 9 /note= "beta-turn"
 FT /note= "D-form residue"
 FT
 XX WO9606629-A1.
 PN 07-MAR-1996.
 XX
 PD 31-AUG-1995; 95WO-US11126.
 XX
 PF 31-AUG-1994; 94US-0299285.
 XX
 XX (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI: 1996-160140/16.
 DR
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Example 2; Page 82; 116pp; English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microM) activities of 51.1 and 4.0, compared to 0.018 and 0.0013
 CC for C5a.
 XX
 SQ Sequence 10 AA:

Query Match 96.0%; Score 48; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.0021;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDWPLXR 10
 I I I I I I I I
 Db 1 ysfkdwplfr 10

RESULT 4
 AAR94468
 ID AAR94468 standard; peptide; 10 AA.
 XX
 AC AAR94468;
 XX
 DT 17-OCT-1996 (first entry)
 XX
 DE C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 Pro71.
 XX
 KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX
 OS Homo sapiens.
 XX
 FH Key "Location/Qualifiers
 FT Region 4..7
 FT /note= "beta-turn"
 FT
 XX WO9606629-A1.
 PN 07-MAR-1996.
 XX
 PD 31-AUG-1995; 95WO-US11126.
 XX
 PF 31-AUG-1994; 94US-0299285.
 XX
 XX (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI: 1996-160140/16.
 DR
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Claim 8; Page 78; 116pp; English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microM) activities of 1.13 and 5, compared to 0.018 and 0.0013
 CC for C5a.
 XX
 SQ Sequence 10 AA:

Query Match 96.0%; Score 48; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.0021;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSFKDMLXR 10
 ||||| |
 Db 1 ysfkdmlgr 10

RESULT 5
 AAR94477
 ID AAR94477 standard; peptide: 10 AA.
 AC AAR94477;
 DT 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 Pro71 D-Pro73.
 DE C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Region 4..7
 FT /note= "beta-turn"
 FT Misc-difference 9
 FT /note= "D-form residue"
 XX WO9606629-AL.
 XX 07-MAR-1996.
 XX 31-AUG-1995; 95WO-US11126.
 XX 31-AUG-1994; 94US-0299285.
 XX (UYNE-) UNIV NEBRASKA.
 XX (UYQU) UNIV QUEENSLAND.
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI: 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX Claim 8; Page 83; 116pp; English.
 XX The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by

CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 3.7 and 2.5, compared to 0.018 and 0.0013
 CC for C5a.
 XX Sequence 10 AA;
 SQ

Query Match 94.0%; Score 47; DB 17; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.0034;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSFKDMLXR 10
 ||||| |
 Db 1 ysfkdmlpr 10

RESULT 6
 AAR94475
 ID AAR94475 standard; peptide: 10 AA.
 XX AAR94475;
 AC AAR94475;
 DT 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 Pro71 Ala72 D-Ala73.
 DE C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Region 4..7
 FT /note= "beta-turn"
 FT Misc-difference 9
 FT /note= "D-form residue"
 XX WO9606629-AL.
 XX 07-MAR-1996.
 XX 31-AUG-1995; 95WO-US11126.
 XX 31-AUG-1994; 94US-0299285.
 XX (UYNE-) UNIV NEBRASKA.
 XX (UYQU) UNIV QUEENSLAND.
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI: 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX Example 2; Page 66; 116pp; English.
 XX The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained

CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 17.9 and 5.6, compared to 0.018 and 0.0013
 CC for C5a.
 XX Sequence 10 AA;
 SQ

Query Match 88.0%; Score 44; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.013;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSPKDMPLXR 10
 I I I I I I I
 Db 1 ysfkdmpaar 10

RESULT 7
 AAR94474
 ID AAR94474 standard; peptide: 10 AA.
 XX AC
 XX AAR94474;
 DT 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 pro71 Ala72.
 KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Region 4..7
 FT /note= "beta-turn"
 FT
 XX WO9606629-A1.
 PN
 XX 07-MAR-1996.
 PD
 XX 31-AUG-1995; 95WO-US11126.
 XX
 PR 31-AUG-1994; 94US-0299285.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 PA (UYQU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX

DR WPI: 1996-160140/16.
 XX
 PT C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Example 2; page 81; 116pp; English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of >1000 and 215, compared to 0.018 and 0.0013
 CC for C5a.
 XX Sequence 10 AA;
 SQ

Query Match 86.0%; Score 43; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.021;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSPKDMPLXR 10
 I I I I I I I
 Db 1 ysfkdmpaar 10

RESULT 8
 AAR94482
 ID AAR94482 standard; peptide: 10 AA.
 XX AC
 XX AAR94482;
 DT 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 Ala70 Pro71 D-Ala73.
 DE
 DE C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 FH Region 4..7
 FT /note= "beta-turn"
 FT Misc-difference /note= "D-form residue"
 FT
 XX WO9606629-A1.
 PN

XX 07-MAR-1996.
 PD XX
 XX 31-AUG-1995; 95WO-US11126.
 PF XX
 XX 31-AUG-1994; 94US-0299285.
 PR XX
 XX (UYNE-) UNIV NEBRASKA.
 PA (UYQU) UNIV QUEENSLAND.
 PA XX
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 PI WPI; 1996-160140/16.
 DR XX
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 PT
 XX Claim 8; Page 85; 116pp; English.
 PS
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 0.33 and 2.3, compared to 0.018 and 0.0013
 CC for C5a.
 CC
 XX Sequence 10 AA;
 SQ

Query Match 86.0%; Score 43; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.021;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YSFKDMPLXR 10
 Db 1 ysfkdaplar 10

RESULT 9
 AAR94478
 ID AAR94478 standard; peptide: 10 AA.
 XX
 AC AAR94478;
 XX
 DT 17-OCT-1996 (first entry)
 XX
 DE C5a anaphylatoxin analogue C5a(65-74) Tyr65 Ala70 Pro71.
 XX
 CC C-terminal: human; C5a anaphylatoxin; analogue: beta-turn;
 CC constrained backbone conformation; proinflammatory response;
 CC spasmogenesis; platelet aggregation; neutrophil: non-mediated;
 CC mediated; increase; cell membrane; vascular; permeability;
 CC neutrophil polarisation; neutrophil enzyme release; treatment;
 CC immune adjuvant; immunodeficiency; augmentation; immune therapy;
 CC cancer; drug delivery; CNS disorder; central nervous system;

KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Region 4..7
 XX /note= "beta-turn"
 XX
 PN WO9606629-A1.
 PD 07-MAR-1996.
 XX
 XX 31-AUG-1995; 95WO-US11126.
 PF XX
 XX 31-AUG-1994; 94US-0299285.
 PR XX
 XX (UYNE-) UNIV NERASKA.
 PA (UYQU) UNIV QUEENSLAND.
 PA XX
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 PI WPI; 1996-160140/16.
 DR XX
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 PT
 XX Claim 8; Page 83; 116pp; English.
 PS
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 1.74 and 15.8, compared to 0.018 and 0.0013
 CC for C5a.
 CC
 XX Sequence 10 AA;
 SQ

Query Match 84.0%; Score 42; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.034;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YSFKDMPLXR 10
 Db 1 ysfkdaplar 10

RESULT 10
 AAR94479
 ID AAR94479 standard; peptide: 10 AA.
 XX
 AC AAR94479;
 XX
 DT 17-OCT-1996 (first entry)

XX DE C5a anaphylatoxin analogue C5a(65-74) Tyr65 Cys70 Pro71.

XX KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;

XX KW constrained backbone conformation; proinflammatory response;

XX KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;

XX KW mediated; increase; cell membrane; vascular; permeability;

XX KW neutrophil polarisation; neutrophil enzyme release; treatment;

XX KW immune adjuvant; immunodeficiency; augmentation; immune therapy;

XX KW cancer; drug delivery; CNS disorder; central nervous system;

XX KW Alzheimer's disease; tumour; high affinity; receptor antagonist;

XX KW development; non-steroidal; anti-inflammatory.

XX OS Homo sapiens.

XX FI Key Location/Qualifiers

XX FT Region 4..7

XX FT /note= "beta-turn"

XX PN W09606629-AL.

XX PD 07-MAR-1996.

XX PF 31-AUG-1995; 95WO-US11126.

XX PF 31-AUG-1994; 94US-0299285.

XX PR (UYNE-) UNIV NEBRASKA.

XX PA (UYQU) UNIV QUEENSLAND.

XX PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;

XX DR WPI: 1996-160140/16.

XX DR C-terminal analogues of C5a anaphylatoxin - induce e.g.

XX PT spasmogenesis, platelet aggregation and increases in cell membrane

XX PT permeability

XX PS Claim 8; Page 84; 116pp; English.

XX CC The present peptide is an analogue of a C-terminal decapeptide

XX CC region from human C5a anaphylatoxin, which has a constrained

XX CC backbone conformation (compared to the natural peptide's flexible

XX CC C-terminal region) comprising a beta-turn. Analogues with type

XX CC (II or V) and type III beta-turns elicit proinflammatory responses

XX CC characterised by spasmogenesis, platelet aggregation and neutrophil

XX CC non-mediated increases of cell membrane (specifically vascular)

XX CC permeability, or neutrophil polarisation, neutrophil enzyme

XX CC release and neutrophil mediated increases in cell membrane (esp.

XX CC vascular) permeability, respectively.

XX CC The analogues can be used as immune adjuvants for the treatment of

XX CC immune deficiency disorders, and for augmenting standard immune

XX CC therapy for the treatment of cancer, without inflammatory side

XX CC effects. They can also be used to facilitate drug delivery by

XX CC increasing vascular permeability, e.g. in the treatment of CNS

XX CC disorders such as Alzheimer's disease or tumours, and to develop

XX CC high affinity C5a receptor antagonists useful as non-steroidal

XX CC anti-inflammatory agents.

XX CC In a type (II or V) beta-turn specific human foetal artery smooth

XX CC muscle contraction assay, and a type III beta-turn specific

XX CC neutrophil polarisation assay the peptide had respective EC50

XX CC (microm) activities of 2.04 and 19.5, compared to 0.018 and 0.0013

XX CC for C5a.

XX SQ Sequence 10 AA;

Query Match 84.0%; Score 42; DB 17; Length 10;

Best Local Similarity 80.0%; Pred. No. 0.034;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YSFKDMPLEXR 10

|||||

Db 1 ysfkdcplgr 10

RESULT 11

AAW27250

ID AAW27250 standard; peptide; 10 AA.

XX AC AAW27250;

XX DT 23-DEC-1997 (first entry)

XX DE Molecular adjuvant targeting ligand.

XX KW Molecular adjuvant; immune response; immunogen; binding affinity;

XX KW antigen presenting cell; APC; viral pathogen; anti tumour response;

XX KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT Misc-difference 9

XX FT /note= "D-form residue"

XX PN W09714426-AL.

XX PD 24-APR-1997.

XX PF 18-OCT-1996; 96WO-US16825.

XX PR 20-OCT-1995; 95US-0005727.

XX PA (UYNE-) UNIV NEBRASKA.

XX PI Hollingsworth MA, Sanderson SD, Tempero RA;

XX DR WPI: 1997-244854/22.

XX PT Molecular adjuvants for enhancing immune responses - comprise an

XX PT immunogen linked to a ligand having binding affinity for a

XX PT determinant of an antigen presenting cell

XX PS Claim 5; Page 51; 61pp; English.

XX CC A novel molecular adjuvant has been developed for enhancing an immune

XX CC response to an immunogen. The molecular adjuvant comprises a targeting

XX CC ligand, having binding affinity for a characteristic determinant of an

XX CC antigen presenting cell (APC), the targeting ligand being functionally

XX CC linked to the immunogen, and so binding of the molecular adjuvant to

XX CC the APC determinant activates the APC, effecting delivery of the

XX CC immunogen to an antigen presenting pathway of the APC. The present

XX CC sequence represents a specifically claimed targeting ligand, which is

XX CC a C5a C-terminal decapeptide agonist. The molecular adjuvant can be used

XX CC to produce an immune response for protecting against viral and other

XX CC pathogens or to produce anti-tumour responses. It can also be used for

XX CC the production of antibodies for use as immunodiagnostic and

XX CC immunotherapeutic agents. The molecular adjuvant is readily taken up and

XX CC processed by APCs to provide APC-mediated immune responses. It can

XX CC provide for the production of antibodies against a weakly-antigenic or

XX CC non-antigenic substances.

XX SQ Sequence 10 AA;

Query Match 84.0%; Score 42; DB 18; Length 10;

Best Local Similarity 80.0%; Pred. No. 0.034;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YSFKDMPLEXR 10

|||||

Db 1 ysfkmpplar 10

RESULT 12

AAW27251
 ID AAW27251 standard: peptide: 19 AA.
 XX AC AAW27251;
 XX DT 23-DEC-1997 (first entry)
 XX DE Molecular adjuvant targeting ligand and immunogen.
 XX KW Molecular adjuvant; immune response; immunogen; binding affinity;
 KW antigen presenting cell; APC; viral pathogen; anti tumour response;
 KW antibody; immunodiagnostic; immunotherapeutic; agonist.
 XX OS Synthetic.
 XX FH Key Location/Qualifiers
 FT Misc-difference 18 /note= "D-form residue"
 XX PN W09714426-A1.
 XX PD 24-APR-1997.
 XX PF 18-OCT-1996: 96WO-US16825.
 XX PR 20-OCT-1995: 95US-0005727.
 XX PA (UYNE-) UNIV NEBRASKA.
 XX PI Hollingsworth MA, Sanderson SD, Tempero RA;
 XX DR WPI: 1997-244854/22.
 XX PT Molecular adjuvants for enhancing immune responses - comprise an
 PT immunogen linked to a ligand having binding affinity for a
 PT determinant of an antigen presenting cell
 XX TS Claim 6: Page 52: 61pp: English.
 XX CC A novel molecular adjuvant has been developed for enhancing an immune
 CC response to an immunogen. The molecular adjuvant comprises a targeting
 CC ligand, having binding affinity for a characteristic determinant of an
 CC antigen presenting cell (APC), the targeting ligand being functionally
 CC linked to the immunogen, and so binding of the molecular adjuvant to
 CC the APC determinant activates the APC, effecting delivery of the
 CC immunogen to an antigen presenting pathway of the APC. The present
 CC sequence represents a specifically claimed targeting ligand and
 CC immunogen. The molecular adjuvant can be used to produce an immune
 CC response for protecting against viral and other pathogens or to
 CC produce anti-tumour responses. It can also be used for the production
 CC of antibodies for use as immunodiagnostic and immunotherapeutic
 CC agents. The molecular adjuvant is readily taken up and processed
 CC by APCs to provide APC-mediated immune responses. It can provide
 CC for the production of antibodies against a weakly-antigenic or
 CC non-antigenic substances.
 XX SQ Sequence 19 AA:
 Query Match 84.0%; Score 42; DB 18; Length 19;
 Best Local Similarity 80.0%; Pred. No. 0.068;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 YSFKDMPLXR 10
 I I I I I I I I
 Db 10 ysfkpmplar 19
 RESULT 15
 AAR94486
 ID AAR94486 standard: peptide: 10 AA.
 XX AC AAR94486;

XX DT 17-OCT-1996 (first entry)
 XX DE C5a anaphylatoxin analogue C5a(65-74) Tyr65 Gly69 Pro70.
 XX KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX OS Homo sapiens.
 XX FH Key Location/Qualifiers
 FT Region 4..7 /note= "beta-turn"
 XX PN W09606629-A1.
 XX PD 07-MAR-1996.
 XX PF 31-AUG-1995: 95WO-US11126.
 XX PR 31-AUG-1994: 94US-0299285.
 XX PA (UYNE-) UNIV NEBRASKA.
 XX PI (UYQU) UNIV QUEENSLAND.
 XX PI Kirnatsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX DR WPI: 1996-160140/16.
 XX PT C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX TS Claim 8: Page 87: 116pp: English.
 XX CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microM) activities of 0.59 and 2.0, compared to 0.018 and 0.0013
 CC for C5a.
 XX SQ Sequence 10 AA:
 Query Match 82.0%; Score 41; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.054;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YSFKDMPXLR 10
 |||| |||
Db 1 ysfkgmpigr 10

Search completed: February 27, 2002, 11:41:13
Job time: 452 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:12 ; Search time 145.23 Seconds
(without alignments)
5.245 Million cell updates/sec

Title: US-09-446-109A-3
Perfect score: 50
Sequence: 1 YSFKDMPLXR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	72.0	295	2 140810	hydroxymethylbilan
2	36	72.0	581	2 G02318	Ets transcription
3	36	72.0	1276	2 T09204	probable tail-host
4	36	72.0	1291	2 T09273	probable tail-host
5	35	70.0	489	2 A81166	type II site-speci
6	34	68.0	555	2 C71414	hypothetical prote
7	34	68.0	1117	2 S33851	fibronectin-bindin
8	33	66.0	271	2 S62485	probable mitochond
9	33	66.0	289	2 A69405	hydroxymethylbilan
10	33	66.0	638	1 YSBS72	threonine--CRNA li
11	33	66.0	1185	2 A42404	collagen adhesin -
12	33	66.0	1469	2 T19459	hypothetical prote
13	33	66.0	1516	2 F83085	conserved hypothet
14	32	64.0	230	2 T24789	hypothetical prote
15	32	64.0	260	2 S71507	restriction endonu
16	32	64.0	305	2 T41372	probable ubiquitin
17	32	64.0	382	2 S58138	gene 8 protein - p
18	32	64.0	576	2 S50590	hypothetical prote
19	32	64.0	593	2 T21510	hypothetical prote
20	32	64.0	615	3 JC7576	transcription fact
21	32	64.0	619	2 A43361	Ets-related transc
22	32	64.0	622	2 E69609	cytochrome-c oxida
23	32	64.0	634	2 T071493	probable DNA helic
24	32	64.0	634	2 C81652	ATP-dependent heli
25	32	64.0	773	2 T01539	hypothetical prote
26	32	64.0	823	2 S18968	cytitestin precurs
27	32	64.0	1019	2 F70342	cation efflux syst
28	32	64.0	1039	2 T30856	protein F2 - Strept
29	32	64.0	1421	2 T49500	hypothetical prote

30 64.0 1738 2 S20614 conserved hypothet
31 64.0 1864 1 JQ1657 genome polyprotein
32 64.0 3229 2 S27852 probable cell-surf
33 62.0 102 2 T44673 H+-transporting AT
34 62.0 131 2 PC4346 phosphoinositide 3
35 62.0 131 2 PC4345 phosphoinositide 3
36 62.0 170 2 D69887 conserved hypothet
37 62.0 201 2 A64029 hypothetrical prote
38 62.0 243 2 T12734 structural protein
39 62.0 265 2 D81315 hypothetrical prote
40 62.0 285 2 F71851 hypothetrical prote
41 62.0 317 2 E69195 conserved hypothet
42 62.0 326 2 A71681 pyruvate dehydroge
43 62.0 378 2 T20266 hypothetrical prote
44 62.0 403 2 A81464 translation initia
45 62.0 440 1 TVVPM middle T antigen -

ALIGNMENTS

RESULT 1

140810
hydroxymethylbilane synthase (EC 4.3.1.8) - Clostridium josui
N:Alternate names: porphobilinogen deaminase
C:Species: Clostridium josui
C>Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C:Accession: 140810
R:Fujino, E.; Fujino, T.; Karita, S.; Sakka, K.; Ohmiya, K.
J. Bacteriol. 177, 5169-5175, 1995
A:Title: Cloning and sequencing of some genes responsible for porphyrin biosynthesis
A:Reference number: A57344; MUID:95394829
A:Accession: 140810
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-295 <RES>
A:Cross-references: GB:D28503; NID:9536874; PIDN:BA05861.1; PID:g460592
C:Genetics:
A:Gene: hemC
C:Superfamily: hydroxymethylbilane synthase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; porphyrin biosynthesis
F:237/Modified site: dipyrrolylmethanemethyl (Cys) (covalent) #status predicted

Query Match 72.0%; Score 36; DB 2; Length 295;
Best local Similarity 75.0%; Pred. No. 7.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSFKDMPL 8
DB 82 HSYKDMPL 89
:|:|:|:|:|

RESULT 2

G02318
Ets transcription factor - human
C:Species: Homo sapiens (man)
C>Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 17-Jul-1998
C:Accession: G02318
R:Liebermann, T.
submitted to the EMBL Data Library, December 1995
A:Reference number: H01044
A:Accession: G02318
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-581 <LIB>
A:Cross-references: EMBL:U43188; NID:g1420888; PID:g1420889
C:Genetics:
A:Gene: NERF-2
C:Superfamily: ets DNA-binding domain homology
F:198-278/Domain: ets DNA-binding domain homology <ETS>

Query Match 72.0%; Score 36; DB 2; Length 581;
 Best Local Similarity 85.7%; Pred. No. 15;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKDMPL 7
 I I I I I
 Db 275 YQFKDMP 281

RESULT 3

T09204
 probable tail-host specificity protein - Streptococcus thermophilus phage Sf121
 C:Species: Streptococcus thermophilus phage Sf121
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
 C:Accession: T09204
 R:Desiere, F.; Lucchini, S.; Brussow, H.
 Virology 241, 345-356, 1998
 A:Title: Evolution of Streptococcus thermophilus bacteriophage genomes by modular exchange
 A:Reference number: 216607; MUID:98160788
 A:Accession: T09204
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1276 <DES>
 A:Cross-references: EMBL:AF032121; NID:g29335667; PID:g29335676
 C:Keywords: tail protein

Query Match 72.0%; Score 36; DB 2; Length 1276;
 Best Local Similarity 75.0%; Pred. No. 36;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKDMPL 8
 I I I I I
 Db 259 YSPKDSMP 266

RESULT 4

T09273
 probable tail-host specificity protein - Streptococcus thermophilus phage Sf119
 C:Species: Streptococcus thermophilus phage Sf119
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
 C:Accession: T09273
 R:Desiere, F.; Lucchini, S.; Brussow, H.
 Virology 241, 345-356, 1998
 A:Title: Evolution of Streptococcus thermophilus bacteriophage genomes by modular exchange
 A:Reference number: 216607; MUID:98160788
 A:Accession: T09273
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1291 <DES>
 A:Cross-references: EMBL:AF032122; NID:g29335682; PID:g29335691

Query Match 72.0%; Score 36; DB 2; Length 1291;
 Best Local Similarity 75.0%; Pred. No. 36;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKDMPL 8
 I I I I I
 Db 259 YSPKDSMP 266

RESULT 5

A81166
 type II site-specific deoxyribonuclease (EC 3.1.21.4) HgaI [imported] - Neisseria meningitidis
 C:Species: Neisseria meningitidis
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
 C:Accession: A81166
 R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
 Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
 ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
 Science 287, 1809-1815, 2000
 A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Va

A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
 A:Reference number: A81000; MUID:20175755
 A:Accession: A81166
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-489 <TET>
 A:Cross-references: GB:AE002428; GB:AE002098; NID:g7225959; PIDN:AAF41139.1; PID:g722
 A:Experimental source: serogroup B, strain MC58
 C:Genetics:

A:Gene: NMB0726

C:Superfamily: Haemophilus paragallinarum type II site-specific deoxyribonuclease Hga
 C:Keywords: hydrolase

Query Match 70.0%; Score 35; DB 2; Length 489;
 Best Local Similarity 75.0%; Pred. No. 20;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKDMPL 8
 I I I I I
 Db 267 YSPSEMP 274

RESULT 6

C71414
 hypothetical protein - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 A:Variety: Columbia
 C:Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 05-Dec-1998
 C:Accession: C71414
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
 avanagh, T.; Hempel, S.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.;
 Nature 391, 485-488, 1998
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomen
 erhoft, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
 C.; Chalvatzis, N.
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
 A:Reference number: A71400; MUID:98121113
 A:Accession: C71414
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-555 <BEV>
 A:Cross-references: GB:297337; NID:g2244829; PID:e326846; PID:g2244863
 C:Genetics:
 A:Map position: 4C09-4C3845

Query Match 68.0%; Score 34; DB 2; Length 555;
 Best Local Similarity 85.7%; Pred. No. 37;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKDMPL 7
 I I I I I
 Db 454 YVFKDMP 460

RESULT 7

S33851
 fibronectin-binding protein precursor - Streptococcus dysgalactiae
 C:Species: Streptococcus dysgalactiae
 C:Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 15-Oct-1999
 C:Accession: S33851; S32631
 R:Lindgren, P.E.; McGavin, M.J.; Signaes, C.; Guss, B.; Gurusiddappa, S.; Hoesek, M.;
 Eur. J. Biochem. 214, 819-827, 1993
 A:Title: Two different genes coding for fibronectin-binding proteins from Streptococ
 A:Reference number: S33850; MUID:93307299
 A:Accession: S33851
 A:Molecule type: DNA
 A:Residues: 1-1117 <LIN>
 A:Cross-references: EMBL:222151; NID:g288970; PIDN:CAA80122.1; PID:g581643
 A:Note: the authors translated the initiation codon TTG for residue 1 as Leu
 C:Genetics:

A:Gene: fnbB
A:Start codon: TTG
C:Keywords: fibronectin binding
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-1117/Product: fibronectin-binding protein #status predicted <MAT>

Query Match 68.0%; Score 34; DB 2; Length 1117;
Best Local Similarity 71.4%; Pred. No. 81;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSFKDMP 7
I:|||||
Db 531 YTFKDLP 537

RESULT 8
S62485
probable mitochondrial carrier - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 16-May-1996 #sequence_revision 13-Mar-1997 #text_change 10-Dec-1999
C:Accession: T38853; S62485
R:Badcock, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
A:Reference number: 221745
A:Accession: T38853
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-271 <BA2>
A:Cross-references: EMBL:Z56276; NID:g1022345; PIDN:CAA91209.1; PID:g1022353; GSPDB:GN000
C:Genetics:
A:Map position: 1L
A:Introns: 10/2; 141/3
A:Superfamily: ADP,ATP carrier protein repeat homology
F:1-75/Domain: ADP,ATP carrier protein repeat homology #status atypical <ACP1>
F:80-164/Domain: ADP,ATP carrier protein repeat homology <ACP2>
F:170-269/Domain: ADP,ATP carrier protein repeat homology <ACP3>

Query Match 66.0%; Score 33; DB 2; Length 271;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMP 8
I:|||||
Db 36 FSKTMPL 43

RESULT 9
A69405
hydroxymethylbilane synthase (EC 4.3.1.8) - Archaeoglobus fulgidus
N:Alternate names: porphobilinogen deaminase
C:Species: Archaeoglobus fulgidus
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
C:Accession: A69405
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon
A:Reference number: A69250; MUID:98049343
A:Accession: A69405
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-289 <KLE>
A:Cross-references: GB:AE000782; NID:g2689341; PIDN:AAB90000.1; PID:g264933
C:Superfamily: hydroxymethylbilane synthase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; porphyrin biosynthesis
F:234/Modified site: dipyrrolylmethanemethyl (Cys) (covalent) #status predicted

Query Match 66.0%; Score 33; DB 2; Length 289;
Best Local Similarity 60.0%; Pred. No. 30;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSFKDMP 10
I:|||||
Db 74 HSKDFVSQR 83

RESULT 10
YSBST2
threonine--tRNA ligase (EC 6.1.1.3) thrZ [validated] - Bacillus subtilis
N:Alternate names: threonine--tRNA ligase, minor; threonyl-tRNA synthetase
C:Species: Bacillus subtilis
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jun-2000
C:Accession: A37770; C69723
R:Putzer, H.; Brakhage, A.A.; Grunberg-Manago, M.
J. Bacteriol. 172, 4593-4602, 1990
A:Title: Independent genes for two threonyl-tRNA synthetases in Bacillus subtilis.
A:Reference number: A37770; MUID:90330571
A:Accession: A37770
A:Molecule type: DNA
A:Residues: 1-638 <PUT>
A:Cross-references: GB:M36593; NID:g143763; PIDN:AAA22863.1; PID:g143764
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmert, P.T.; Entian, K.D.; Erlington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Iijima, M.
Koetter, P.; Koningsstein, G.; Krogh, H.; Kurita, K.; Kumano, M.; Kurita, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Porleto
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69580; MUID:98044033
A:Accession: C69723
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-638 <KUN>
A:Cross-references: GB:Z99123; GB:AL009126; NID:g2636240; PIDN:CAB15783.1; PID:g26362
A:Experimental source: strain 168
C:Comment: This threonine--tRNA ligase is not expressed during vegetative growth.
C:Genetics:
A:Gene: thrZ; thrS2
A:Map position: 344 (degrees)
C:Function:
A:Description: EC 6.1.1.3 [validated; MUID:90330571]
C:Superfamily: threonine--tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ATP; ligase; protein biosynthesis

Query Match 66.0%; Score 33; DB 1; Length 638;
Best Local Similarity 50.0%; Pred. No. 71;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSFKDMP 8
I:|||||
Db 348 YSRDLP 355

RESULT 11
A42404
collagen adhesin - Staphylococcus aureus
C:Species: Staphylococcus aureus
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Nov-1995
C:Accession: A42404; S27665
R:Patti, J.M.; Jonsson, H.; Guss, B.; Switalski, L.M.; Wiberg, K.; Lindberg, M.; Hook
J. Biol. Chem. 267, 4766-4772, 1992

A:Title: Molecular characterization and expression of a gene encoding a Staphylococcus

Best Local Similarity 60.0%; Pred. No. 1.8e+02; Mismatches 6; Conservative 1; Mismatches 3; Indels 0; Caps 0;

Matches 6; Conservative 1; Mismatches 3; Indels 0; Caps 0;

QY 1 YSFKDMP 10
I: I I I I I

DB 334 YAFYDRPLR 343

RESULT 14
T24789

hypothetical protein T10C6.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T24789

R:McMurray, A.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19936

A:Accession: T24789

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-230 <WIL>

A:Cross-references: EMBL:Z93388; PIDN:CAB07655.1; GSPDB:GN00023; CESP:T10C6.5

A:Experimental source: clone T10C6

C:Genetics:

A:Gene: CESP:T10C6.5

A:Map position: 5

A:Introns: 57/3

Query Match 64.0%; Score 32; DB 2; Length 230;

Best Local Similarity 85.7%; Pred. No. 37;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 YSFKDMP 7
I: I I I I I

DB 30 YSSKDMP 36

RESULT 15
S71507

restriction endonuclease (EC 3.1.1.-) BslI - Brevibacterium albidum (ATCC 15831)

C:Species: Brevibacterium albidum

A:Variety: ATCC 15831

C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 20-Jun-2000

C:Accession: S71507

R:Ueno, H.; Kato, I.; Ishino, Y.

Nucleic Acids Res. 24, 2268-2270, 1996

A:Title: Cloning and expression of the BslI restriction-modification system.

A:Reference number: S71506; MUID:96279725

A:Accession: S71507

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-260 <UEN>

A:Cross-references: EMBL:D82028; NID:g1620875; PIDN:BAAL1515.1; PID:g1620877

A:Experimental source: ATCC 15831

A:Note: the amino end of this sequence was confirmed by protein sequencing

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1996

C:Genetics:

A:Start codon: GTG

C:Superfamily: Brevibacterium albidum restriction endonuclease BslI

C:Keywords: endonuclease; hydrolase; methyltransferase; restriction modification sys

Query Match 64.0%; Score 32; DB 2; Length 260;

Best Local Similarity 62.5%; Pred. No. 43;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Caps 0;

QY 1 YSFKDMP 8
I: I I I I I

DB 3 YAFDRPL 10

A:Title: Molecular characterization and expression of a gene encoding a Staphylococcus

A:Reference number: A42404; MUID:92165839

A:Contents: FDA 574

A:Accession: A42404

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1185 <PAT>

A:Cross-references: EMBL:M81736

A:Note: sequence extracted from NCBI backbone (NCBIP:83982)

Query Match 66.0%; Score 33; DB 2; Length 1185;

Best Local Similarity 71.4%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Caps 0;

QY 1 YSFKDMP 7
I: I I I I I

DB 583 YEFKDL 589

RESULT 12
T19459

hypothetical protein C25F9.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19459

R:Mortimore, B.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19126

A:Accession: T19459

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1469 <WIL>

A:Cross-references: EMBL:Z81476; PIDN:CAB03918.1; GSPDB:GN00023; CESP:C25F9.2

A:Experimental source: clone C25F9

C:Genetics:

A:Gene: CESP:C25F9.2

A:Map position: 5

A:Introns: 119/1; 232/2; 315/3; 369/3; 464/2; 535/2; 612/1; 864/3; 965/3; 1038/3; 1113/1

Query Match 66.0%; Score 33; DB 2; Length 1469;

Best Local Similarity 71.4%; Pred. No. 1.8e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Caps 0;

QY 1 YSFKDMP 7
I: I I I I I

DB 649 YSYKDVP 655

RESULT 13
F83085

conserved hypothetical protein PA4489 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83085

K:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337

A:Accession: F83085

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1516 <STO>

A:Cross-references: GB:AE004862; GB:AE004091; NID:g9950716; PIDN:AAG07877.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4489

Query Match 66.0%; Score 33; DB 2; Length 1516;

Search completed: February 27, 2002, 11:45:13
Job time: 692 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:39 ; Search time 78.39 Seconds
(without alignments)
4.677 Million cell updates/sec

Title: US-09-446-109A-3
Perfect score: 50
Sequence: 1 YSKDMLXR 10

Scoring table: BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	72.0	295	1 HEM3_CLOJO	Q59293 clostridium
2	33	66.0	271	1 YAD8_SCHPO	Q09834 schizosacch
3	33	66.0	289	1 HEM3_ARCFU	Q29026 archaeoglob
4	33	66.0	521	1 RF3_STAAU	Q86490 staphylococ
5	33	66.0	638	1 SYT2_BACSU	P18256 bacillus su
6	33	66.0	1183	1 CNA_STAAU	Q53654 staphylococ
7	32	64.0	295	1 MAUJ_METEX	Q49127 methylobact
8	32	64.0	576	1 SYPC_YEAST	P39965 saccharomyc
9	32	64.0	612	1 ELFI_MOUSE	Q60775 mus musculu
10	32	64.0	619	1 ELFI_HUMAN	P32519 homo sapien
11	32	64.0	622	1 COX1_BACSU	P24010 bacillus su
12	32	64.0	1738	1 YCF1_EPTVI	Q00383 epifagus vi
13	32	64.0	1864	1 VGNB_RCAV	P35930 red clover
14	31	62.0	102	1 VATF_DESSY	Q06503 desulfuroco
15	31	62.0	201	1 YE18_HAEIN	P44189 haemophilus
16	31	62.0	326	1 ODP4_RICPR	Q9zdr4 rickettsia
17	31	62.0	355	1 TRUB_THEAC	Q9h1c4 thermoplas
18	31	62.0	403	1 IF4A_LEIRR	Q25225 leishmania
19	31	62.0	440	1 TAMI_POVW3	P03076 mouse polyo
20	31	62.0	567	1 N02M_HANW1	P48906 hansenuita w
21	31	62.0	719	1 P5CS_MESCR	Q65361 m delta l-p
22	31	62.0	1139	1 MSH5_CAEEL	Q19272 caenorhabd
23	31	62.0	1588	1 AR01_YEAST	P08566 s pentafunc
24	31	62.0	1634	1 PK3B_HUMAN	Q00750 homo sapien
25	30	60.0	82	1 YFAE_HAEIN	P45154 haemophilus
26	30	60.0	134	1 CYB_CENSE	Q34253 centurio s
27	30	60.0	143	1 Y4AM_RHISN	P55360 rhizobium s
28	30	60.0	232	1 UPP_PYRAB	Q9v0k1 pyrococcus
29	30	60.0	276	1 Y009_METJA	Q60320 methanococ
30	30	60.0	332	1 GALT_R_HAEIN	P31766 haemophilus
31	30	60.0	356	1 VATC_THEAC	Q9hm67 thermoplas
32	30	60.0	396	1 YWBD_BACSU	P39587 bacillus su
33	30	60.0	403	1 HMPA_ALCEU	P39662 alcaligenes

ALIGNMENTS

```

RESULT 1
HEM3_CLOJO
ID HEM3_CLOJO STANDARD: PRT: 295 AA.
AC Q59293;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PORPHOBILINOGEN DEAMINASE (EC 4.3.1.8) (PBG) (HYDROXYMETHYLBILANE
DE SYNTHASE) (HMB) (PRE-UROPORPHYRINOGEN SYNTHASE).
DE HEMC.
GN Clostridium josui.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium
OX NCBI_TaxID=1499;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FERM P-9684;
RX MEDLINE=95394829; PubMed=7665501;
RA Fujino E., Fujino T., Karita S., Sakka K., Ohmiya K.:
RT "Cloning and sequencing of some genes responsible for porphyrin
RT biosynthesis from the anaerobic bacterium Clostridium josui.";
RL J. Bacteriol. 177:5169-5175(1995).
CC -!- FUNCTION: TETRAPOLYMERIZATION OF THE MONOPYRROLE PBG INTO THE
CC HYDROXYMETHYLBILANE PREUROPORPHYRINOGEN + H(2)O =
CC -!- CATALYTIC ACTIVITY: 4 PORPHOBILINOGEN + 4 NH(3) =
CC HYDROXYMETHYLBILANE + 4 NH(3)
CC -!- COFACTOR: COVALENTLY BINDS A DTPYRROMETHANE COFACTOR TO WHICH
CC THE PORPHOBILINOGEN SUBUNITS ARE ADDED.
CC -!- PATHWAY: STROHEME BIOSYNTHESIS
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE HMB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D28503; BAA05861.1; -
CC HSP: P06983; IPDA
CC InterPro: IPR000860; Porphobil_deam.
CC Pfam: PF01379; Porphobil_deam; 1.
CC PRINTS: PR00151; PORPHOBDMNASE.
CC PROSITE: PS00533; PORPHOBILINOGEN_DEAM; 1.
CC Porphyrin biosynthesis; Lyase.
CC BINDING 237 237
CC SEQUENCE 295 AA; 32744 MW; C0037415E92AE50 CRC64;
DR HSP: P06983; IPDA
DR InterPro: IPR000860; Porphobil_deam.
DR Pfam: PF01379; Porphobil_deam; 1.
DR PRINTS: PR00151; PORPHOBDMNASE.
DR PROSITE: PS00533; PORPHOBILINOGEN_DEAM; 1.
KW Porphyrin biosynthesis; Lyase.
FT BINDING 237 237
SQ SEQUENCE 295 AA; 32744 MW; C0037415E92AE50 CRC64;

```

Query Match 72.0% Score 36; DB 1; Length 295;
Best Local Similarity 75.0%; Pred. No. 3;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Cy 1 YSKDML 8

```

Db      82 HSYKDMPL 89
      1:|||||
RESULT 2
YAD8_SCHPO STANDARD; PRT; 271 AA.
ID YAD8_SCHPO
AC Q09834;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE PUTATIVE-MITOCHONDRIAL CARRIER C4G8.08.
GN SPAC4G8.08.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Badcock K., Churcher C.M., Bartell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (POTENTIAL).
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; Z56276; CAA91209.1;
DR InterPro: IPR001993; Mitoch_carrier.
DR Pfam: PF00153; mito_carr; 3
DR PROSITE: PS00215; MITOCH_CARRIER; 1.
KW Hypothetical protein; Mitochondrion; Inner membrane; Repeat;
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 84 104 POTENTIAL.
FT TRANSMEM 138 158 POTENTIAL.
FT TRANSMEM 170 190 POTENTIAL.
SQ SEQUENCE 271 AA; 30002 MW; EF330C8CC8BDB6C4 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 271;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKDMPL 8
   1:|||||
Db 36 FSEKTMPL 43

RESULT 3
HEM3_ARCFU STANDARD; PRT; 289 AA.
ID HEM3_ARCFU
AC Q29026;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE PORPHOBILINOGEN DEAMINASE (EC 4.3.1.8) (PBG)
DE (HYDROXYMETHYLBILANE SYNTHASE) (HMBS) (PRE-UPROPORPHYRINOGEN SYNTHASE).
GN HEMC OR AF1242
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]

SEQUENCE FROM N.A.
STRAIN=VC-16 / DSM 4304 / ATCC 49558;
MEDLINE=98049343; PubMed=9389475;
Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
Overbeek R., Cocayne J.D., Weidman J.F., McAnold L., Utterback T.,
Cotton M.D., Spriggs T., Artlach P., Kaine B.P., Sykes S.M.,
Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
Venter J.C.;
*The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.*;
RL Nature 390:364-370(1997).
CC -1- FUNCTION: TETRAPOLYMERIZATION OF THE MONOPYRROLE PBG INTO THE
CC HYDROXYMETHYLBILANE PREUROPORPHYRINOGEN IN SEVERAL DISCRETE STEPS
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 4 PORPHOBILINOGEN + H(2)O -
CC HYDROXYMETHYLBILANE + 4 NH(3).
CC -1- COFACTOR: COVALENTLY BINDS A DIPHYROMETHANE COFACTOR TO WHICH
CC THE PORPHOBILINOGEN SUBUNITS ARE ADDED (BY SIMILARITY).
CC -1- PATHWAY: FOURTH STEP IN PORPHYRIN BIOSYNTHESIS BY THE C5 PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE HMBS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; AF001018; AAB90000.1;
DR TIGR; AF1242;
DR InterPro: IPR000860; Porphobil_deam.
DR Pfam: PF01379; Porphobil_deam; 1.
DR PRINTS: PR00151; PORPHBDMNASE.
DR PROSITE: PS00533; PORPHOBILINOGEN_DEAM; FALSE_NEG.
KW Porphyrin biosynthesis; Lyase; Complete proteome.
FT BINDING 234 234 PYRROMETHANE COFACTOR (BY SIMILARITY).
SQ SEQUENCE 289 AA; 32006 MW; E3527A0F9C09C78C CRC64;

Query Match 66.0%; Score 33; DB 1; Length 289;
Best Local Similarity 60.0%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSPKDMPLXR 10
   1:|||||
Db 74 HSPKDVPSQR 83

RESULT 4
RF3_STAAU STANDARD; PRT; 521 AA.
ID RF3_STAAU
AC Q86490;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PEPTIDE CHAIN RELEASE FACTOR 3 (RF-3).
GN PRFC OR RF3.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RX MEDLINE=98313013; PubMed=9650993;
RA Ludovice A.M., Wu S., de Lencastre H.;

```

*Molecular cloning and DNA sequencing of the Staphylococcus aureus
 RT UDP-N-acetylmuramyl tripeptide synthetase (murE) gene, essential for
 RT the optimal expression of methicillin resistance.;
 RL Microb. Drug Resist. 4:85-90(1998).
 CC -1- FUNCTION: INCREASES THE FORMATION OF RIBOSOMAL TERMINATION
 CC COMPLEXES AND STIMULATES ACTIVITIES OF RF-1 AND RF-2. IT BINDS
 CC GUANINE NUCLEOTIDES AND HAS STRONG PREFERENCE FOR UGA STOP CODONS.
 CC IT MAY INTERACT DIRECTLY WITH THE RIBOSOME. THE STIMULATION OF RF-
 CC 1 AND RF-2 IS SIGNIFICANTLY REDUCED BY GTP AND GDP, BUT NOT BY
 CC GMP (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC PRCF SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y14370; CAA74739.1; -
 CC InterPro: IPR000795; GTP_EFTU.
 CC Pfam; PF00009; GTP_EFTU; 1.
 CC PROSITE; PS00301; EFATOR_GTP; FALSE_NEG.
 CC KW Protein biosynthesis; GTP-binding.
 CC FT NP_BIND 17 24 GTP (BY SIMILARITY).
 CC FT NP_BIND 85 89 GTP (BY SIMILARITY).
 CC FT NP_BIND 139 142 GTP (BY SIMILARITY).
 CC SQ SEQUENCE 521 AA; 60016 MW; C39DBDB3F4BBE460 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 521;
 Best Local Similarity 71.4%; Pred. No. 24;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 OY 1 YSPKDMPL 7
 ||:|:|
 Db 380 YSFODLP 386

RESULT 5
 SYT2_BACSU STANDARD; PRT; 638 AA.
 AC P18256; P70992;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE THREONYL-TRNA SYNTHETASE 2 (EC 6.1.1.3) (THREONINE--TRNA LIGASE)
 (THRS).
 GN THRS2
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90330571; PubMed=2115870;
 RA Putzer H., Brakhage A., Grunberg-Manago M.;
 RT "Independent genes for two threonyl-tRNA synthetases in Bacillus
 subtilis.";
 RL J. Bacteriol. 172:4593-4602(1990).
 RN [2]
 RP SEQUENCE OF 38-638 FROM N.A.
 RC STRAIN=168;
 RA Glaser P., Sekowska A., Danchin A.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP EXPRESSION.
 RX MEDLINE=92347349; PubMed=1379177;
 RA Putzer H., Gendron N., Grunberg-Manago M.;
 RT "Co-ordinate expression of the two threonyl-tRNA synthetase genes in

Bacillus subtilis: control by transcriptional antitermination
 RT involving a conserved regulatory sequence.";
 RL EMBO J. 11:3117-3127(1992).
 CC -1- CATALYTIC ACTIVITY: ATP + L-THREONINE + TRNA(THR) -> AMP +
 CC PYROPHOSPHATE + L-THREONYL-TRNA(THR).
 CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- DEVELOPMENTAL STAGE: NORMALLY NOT EXPRESSED. ITS EXPRESSION IS
 CC INDUCED WHEN THAT OF TRS IS REDUCED.
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M36593; AAA22863.1; -
 CC EMBL; Z80360; CAB02510.1; -
 CC EMBL; Z99123; CAB15783.1; -
 CC PIR; A37770; YSBST2.
 CC Subtilist; BCL0421; thrZ.
 CC InterPro: IPR002106; AA_trna_ligase_II.
 CC InterPro: IPR002314; tRNA-synt_2b.
 CC InterPro: IPR002320; tRNA-synt_thr.
 CC Pfam; PF00587; tRNA-synt_2b; 1.
 CC PRINTS; PR01047; TRNASYNTHTHR.
 CC PROSITE; PS00179; AA-trna_ligase_II_1; 1.
 CC PROSITE; PS00339; AA-trna_ligase_II_2; 1.
 CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Metal-binding; Zinc; Multigene family; Complete proteome.
 FT DOMAIN 245 535 CATALYTIC.
 FT METAL 336 336 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 387 387 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 512 512 ZINC (CATALYTIC) (BY SIMILARITY).
 FT CONFLICT 74 74 V -> L (IN REF. 2).
 SQ SEQUENCE 638 AA; 73373 MW; D61922564598975A CRC64;

Query Match 66.0%; Score 33; DB 1; Length 638;
 Best Local Similarity 50.0%; Pred. No. 29;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
 OY 1 YSPKDMPL 8
 ||:|:|
 Db 348 YSKDLPI 355

RESULT 6
 CNA_STAAU STANDARD; PRT; 1183 AA.
 ID CNA_STAAU
 AC Q53634;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE COLLAGEN ADHESIN PRECURSOR.
 GN CNA.
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=1280;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FDA 574;
 RX MEDLINE=92165819; PubMed=1311320;
 RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
 RA Lindberg M., Honorek M.;
 RT "Molecular characterization and expression of a gene encoding a
 RT Staphylococcus aureus collagen adhesin.";
 RL J. Biol. Chem. 267:4766-4772(1992).

```

RN RA Patti J.M., Jonsson H., Cuss B., Switalski L.M., Wiberg K.,
RP Lindberg M., Hooeok M.;
RL J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hooeok M.;
RT Identification and biochemical characterization of the ligand
RL binding domain of the collagen adhesin from Staphylococcus aureus.;
RN Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Symersky J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., DeLucas L.J., Hooeok M.,
RA Narayana S.V.L.;
RT Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin.;
RL Nat. Struct. Biol. 4:833-838(1997).
CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYLOCOCCAL CELLS TO
CC COLLAGEN-CONTAINING SUBSTRATA.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. CELL WALL.
CC -1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M81736; AAA20874.1;
DR PDB: LAMX; 24-JUN-98.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; FALSE_NEG.
KW Signal; Repeat; Transmembrane; Cell wall; 3D-structure.
FT SIGNAL 1 29
FT CHAIN 30 1183 COLLAGEN ADHESIN.
FT DOMAIN 30 1157 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1158 1177 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 1178 1183 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 151 318 COLLAGEN-BINDING.
FT DOMAIN 533 1093 3 X 187 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 1093 1157 LYS/PRO-RICH (CELL WALL-SPANNING).
FT DOMAIN 1151 1156 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT REPEAT 533 719 B1.
FT REPEAT 720 906 B2.
FT REPEAT 907 1093 B3.
SQ SEQUENCE 1183 AA; 133066 MW; B6ALCC072E575D76 CRC64;

Query Match 66.0%; Score 33; DB 1; Length 1183;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMF 7
I I I I I
Db 581 YERKDLF 587

RESULT 7
ID MAUJ_METEX STANDARD; PRT; 295 AA.
AC Q49127;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)

```

```

DE METHYLAMINE UTILIZATION PROTEIN MAUJ.
GN MAUJ.
OS Methylobacterium extorquens.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Methylobacterium group; Methylobacterium.
OX NCBI_TaxID=408;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AM1 / NCIB 9133;
RX MEDLINE=94292425; PubMed=8021187;
RA Chistoserdov A.Y., Chistoserdova L.V., McIntire W.S., Lidstrom M.E.;
RT "Genetic organization of the mau gene cluster in Methylobacterium
RT extorquens AM1: complete nucleotide sequence and genotation and
RT characteristics of mau mutants";
RL J. Bacteriol. 176:4052-4065(1994).
CC -1- PATHWAY: METHYLAMINE UTILIZATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L26406; AAB46938.1;
DR SEQUENCE 295 AA; 32682 MW; 7FE490E025D99191 CRC64;

Query Match 64.0%; Score 32; DB 1; Length 295;
Best Local Similarity 85.7%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 SFKDMPL 8
I I I I I
Db 222 SFKQDPL 228

RESULT 8
SYNCP_YEAST
ID SYNCP_YEAST STANDARD; PRT; 576 AA.
AC P39965;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PROBABLE POLYL-TRNA SYNTHETASE, CYTOPLASMIC (EC 6.1.1.15) (PROLINE--
DE TRNA LIGASE) (PRORS).
GN YER087W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berne A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Bolstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-PROLINE + TRNA(Pro) -> AMP +
CC PYROPHOSPHATE + L-PROLYL-TRNA(Pro).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

```


DR HSP: P14921; 2STW.
 DR TRANSFAC: T01113;
 DR MW: 189973;
 DR InterPro: IPR000418; Ets.
 DR InterPro: IPR002341; HSF_ETs.
 DR Pfam: PF00178; Ets; 1.
 DR PRINTS: PR00454; ETSDOMAIN.
 DR SMART: SM00413; ETS; 1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS00061; ETS_DOMAIN_3; 1.
 DR Nuclear protein: Transcription regulation; Activator; DNA-binding.
 KW DOMAIN: 75-80 POLY-ASP.
 FT DNA_BIND: 208 290 ETS-DOMAIN.
 FT CONFLICT: 283 283 Q -> G (IN REF. 2).
 SO SEQUENCE 619 AA; 67455 MW; AB0841B2964A66EF CRC64;

Query Match 64.0%; Score 32; DB 1; Length 619;
 Best Local Similarity 71.4%; Pred. No. 46;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSEKDMF 7
 I I I I I
 Db 287 YQKEMP 293

RESULT 11
 COXI_BACSU STANDARD: PRT: 622 AA.
 AC P24010: Q34467;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1) (CYTOCHROME AA3
 DE SUBUNIT 1) (CAA-3605 SUBUNIT 1).
 GN CTAD.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=168;
 RX MEDLINE=91146590; PubMed=1847686;
 RA Saraste M., Metsu T., Nakari T., Jalli T., Lauraeus M.,
 van der Oost J.;
 RT "The Bacillus subtilis cytochrome-c oxidase. Variations on a
 conserved protein theme";
 RL Eur. J. Biochem. 195;517-525(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Bertero M., Presecan E., Glaser P., Richou A., Danchin A.;
 RT "Bacillus subtilis chromosomal region downstream nprE";
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 CC AND COPPER B. THIS CYTOCHROME C OXIDASE SHOWS PROTON PUMP ACTIVITY
 CC ACROSS THE MEMBRANE IN ADDITION TO THE ELECTRON TRANSFER.
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) -> 2 H(2)O +
 CC 4 FERRICYTOCHROME C.
 CC -1- COFACTOR: TWO HEME GROUPS AND COPPER B.
 CC -1- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
 CC -1- SURCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CONTAINS 14
 CC POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54140; CAA38077.1; -
 DR EMBL: Z98682; CAB11343.1; -
 DR EMBL: Z99111; CAB13363.1; -
 DR PIR: S14397; S14397.
 DR HSSP: P98002; IAR1.
 DR Subtilist: BG10216; ctad.
 DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1; 1.
 DR PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain;
 KW Hydrogen ion transport; Complete proteome.
 FT DOMAIN 1 27
 FT TRANSMEM 28 46
 FT DOMAIN 47 68
 FT TRANSMEM 69 88
 FT DOMAIN 89 110
 FT TRANSMEM 111 128
 FT DOMAIN 129 159
 FT TRANSMEM 160 178
 FT DOMAIN 179 196
 FT TRANSMEM 197 215
 FT DOMAIN 216 241
 FT TRANSMEM 242 261
 FT DOMAIN 262 284
 FT TRANSMEM 285 304
 FT DOMAIN 305 312
 FT TRANSMEM 313 331
 FT DOMAIN 332 346
 FT TRANSMEM 347 366
 FT DOMAIN 367 374
 FT TRANSMEM 375 394
 FT DOMAIN 395 421
 FT TRANSMEM 422 441
 FT DOMAIN 442 459
 FT TRANSMEM 460 479
 FT DOMAIN 480 552
 FT TRANSMEM 553 572
 FT DOMAIN 573 580
 FT TRANSMEM 581 604
 FT DOMAIN 605 622
 FT METAL 73 73
 FT METAL 249 249
 FT METAL 253 253
 FT METAL 298 298
 FT METAL 299 299
 FT METAL 384 384
 FT METAL 386 386
 FT CONFLICT 120 120
 FT CONFLICT 155 156
 FT CONFLICT 288 288
 FT CONFLICT 474 474
 SQ SEQUENCE 622 AA; 69107 MW; 80159F21D1913068 CRC64;

Query Match 64.0%; Score 32; DB 1; Length 622;
 Best Local Similarity 50.0%; Pred. No. 47;
 Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSEKDMPLXR 10
 I I I I I
 Db 514 YNFKQLPFVR 523

RESULT 12
 YCFL_EPVI

ID YCF1_EPIV1 STANDARD: PRT: 1738 AA.
AC Q00383;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 208 KDA PROTEIN YCF1 (ORF 1738).
GN YCF1.
OS Epitagus virginiana (Beechdrops).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Lamiales; Orobanchaceae; Epifagus.
OX NCBI_TaxID=4177;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92114084; PubMed=1731088;
RA Wolfe K.H., Morden C.W., Palmer J.D.;
RT "Small single-copy region of plastid DNA in the non-photosynthetic angiosperm Epifagus virginiana contains only two genes. Differences among dicots, monocots and bryophytes in gene organization at a non-bioenergetic locus.";
RT J. Mol. Biol. 223:95-104(1992).
RL [2]
RN SEQUENCE FROM N.A.
RP MEDLINE=93066301; PubMed=1332054;
RX Wolfe K.H., Morden C.W., Palmer J.D.;
RA "Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant.";
RT Proc. Natl. Acad. Sci. U.S.A. 89:10648-10652(1992).
RL CC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS TO THE YCF1 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC ENBL: X61368; CAA43644.1; -;
DR ENBL: M81884; AAA65870.1; -;
DR PIR: S16720; S16720.
DR Mendel: S20614; S20614.
DR Mendel: S139; EPIV1; ycf1.1.
KW Chloroplast; Hypothetical protein.
SO SEQUENCE 1738 AA; 208318 MW; 65C63F63BDC8364B CRC64;

Query Match 64.0%; Score 32; DB 1; Length 1738;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 YSFKMDPL 8
| | | | |
Db 997 YSFKMDL 1004

RESULT 13
VGNB_RCMV STANDARD: PRT: 1864 AA.
AC P35930;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GENOME POLYPROTEIN B (CONTAINS: PROTEASE COFACTOR; MEMBRANE BINDING PROTEIN; VPG; PROTEASE (EC 3.4.22.-); RNA POLYMERASE (EC 2.7.7.48)).
OS Red clover mottle virus (RCMV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Comovirus.
OX NCBI_TaxID=12262;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-S;
RX MEDLINE=93019077; PubMed=1402822;
RA Shanks M., Lomonosoff G.P.;
RT "The nucleotide sequence of red clover mottle virus bottom component RNA.";
RL J. Gen. Virol. 73:2473-2477(1992).
CC -1- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -1- SIMILARITY: TO OTHER COMOVIRUSES GENOME POLYPROTEIN B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC ENBL: X64886; CAA46104.1; -;
DR PIR: JQ1657; JQ1657.
DR MEROPS: C03.003; -;
DR InterPro: IPR000605; RNA_helicase.
DR InterPro: IPR001205; RNA_pol_P3D.
DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
DR Pfam: PF00910; RNA_helicase; 1.
KW Polypeptide; Transmembrane; Hydrolase; Protease; Transferase;
KW RNA-directed RNA polymerase; ATP-binding
FT CHAIN 1 315 PROTEASE COFACTOR (POTENTIAL).
FT CHAIN 316 915 MEMBRANE-BINDING PROTEIN (POTENTIAL).
FT CHAIN 916 943 VPG PROTEIN (POTENTIAL).
FT CHAIN 944 1151 PROTEASE (POTENTIAL).
FT CHAIN 1152 1864 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).
FT NP_BIND 483 490 ATP (POTENTIAL).
FT ACT_SITE 983 983 THIOL PROTEASE (POTENTIAL).
FT ACT_SITE 1019 1019 THIOL PROTEASE (POTENTIAL).
FT ACT_SITE 1109 1109 THIOL PROTEASE (POTENTIAL).
SQ SEQUENCE 1864 AA; 210255 MW; 8DDDD9D6C17C2B154 CRC64;

Query Match 64.0%; Score 32; DB 1; Length 1864;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
QY 1 YSFKMDPLXR 10
| : : | | |
Db 927 YRYKNVPLTR 936

RESULT 14
VATF_DESSY STANDARD: PRT: 102 AA.
AC O06503;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE V-TYPE ATP SYNTHASE SUBUNIT F (EC 3.6.1.34) (V-TYPE ATPASE SUBUNIT F).
GN ATPF.
OS Desulfurococcus sp. (strain SY).
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OC Desulfurococcus.
OX NCBI_TaxID=59822;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97320421; PubMed=9177272;
RA Shibui H., Hamamoto T., Yoshida M., Kagawa Y.;
RT "The stabilizing residues and the functional domains in the hyperthermophilic V-ATPase of Desulfurococcus";
RL Biochem. Biophys. Res. Commun. 234:341-345(1997).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON GRADIENT ACROSS THE MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE V-ATPASE F SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U96487; AAB64415.1; ..
DR InterPro: IPR002841; ATP-synt_F.
DR Pfam: PF01990; ATP-synt_F; 1.
DR ProDom: PD003811; ATP-synt_F; 1.
KW Hydrolase; ATP synthetase; Hydrogen ion transport.
SQ SEQUENCE 102 AA: 11499 MW: 480DE0FBFDADF5B CRC64;

Query Match 62.0% Score 31; DB 1; Length 102;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Caps 0;
QY 1 YSPKDMPL 8
I:I:I I I
DB 24 YAFEDTPL 31

RESULT 15
YE18_HAFIN
ID YE18_HAFIN STANDARD; PRT; 201 AA.
AC P44189;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN H11418.
GN H11418.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: U32821; AAC23068.1; ..
DR TIGR: H11418; ..
DR InterPro: IPR003497; BRO.
DR Pfam: PF02498; BRO; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 201 AA: 23567 MW: 87964F8E43B4327B CRC64;

Query Match 62.0% Score 31; DB 1; Length 201;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Caps 0;

QY 1 YSPKDMPL 8
DB 23 FNFKDLPV 30

Search completed: February 27, 2002, 11:42:41
Job time: 540 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:01 ; Search time 281.76 Seconds
(without alignments)
5.191 Million cell updates/sec

Title: US-09-446-109A-3
Perfect score: -50
Sequence: 1 YSKDMLXR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum-DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	76.0	1492	10 Q9FHA3	Q9fha3 arabidopsis
2	36	72.0	521	4 Q15725	Q15725 homo sapien
3	36	72.0	521	11 Q5JHC8	Q5jhc8 mus musculus
4	36	72.0	533	4 Q15724	Q15724 homo sapien
5	36	72.0	533	11 Q5JHC7	Q5jhc7 mus musculus
6	36	72.0	581	4 Q15723	Q15723 homo sapien
7	36	72.0	581	11 Q9JHD0	Q9jhd0 mus musculus
8	36	72.0	593	11 Q9JHC9	Q9jhc9 mus musculus
9	36	72.0	914	9 Q9XJ81	Q9xjb1 streptococc
10	36	72.0	1276	9 Q64284	Q64284 streptococc
11	36	72.0	1291	9 Q64295	Q64295 streptococc
12	35	70.0	175	2 Q92611	Q92611 borrelia tu
13	35	70.0	473	2 Q9XAC6	Q9xaz6 neisseria m
14	35	70.0	489	2 Q9K091	Q9k091 neisseria m
15	34	68.0	247	11 Q9DAG5	Q9dag5 mus musculus
16	34	68.0	555	10 Q23360	Q23360 arabidopsis
17	34	68.0	566	2 P71093	P71093 bacteroides
18	34	68.0	1117	2 Q53971	Q53971 streptococc
19	34	68.0	1579	5 Q9GRT7	Q9grt7 leishmania

20	33	66.0	206	2	Q99VH2	Q99vh2 staphylococ
21	33	66.0	373	11	Q92311	Q92311 rattus norv
22	33	66.0	373	11	Q9DCS3	Q9dcs3 mus musculu
23	33	66.0	373	11	Q99L39	Q99l39 mus musculu
24	33	66.0	407	2	Q9A5G9	Q9a5g9 caulobacter
25	33	66.0	520	2	Q99V72	Q99v72 staphylococ
26	33	66.0	546	10	P93762	P93762 chondrus cr
27	33	66.0	619	2	Q45531	Q45531 bacillus sp
28	33	66.0	750	2	P70921	P70921 bradyrhizob
29	33	66.0	1151	3	Q9HEF9	Q9hef9 neurospora
30	33	66.0	1469	5	Q62074	Q62074 caenorhabdi
31	33	66.0	1516	2	Q9HVT2	Q9hvt2 pseudomonas
32	33	66.0	3070	12	Q89906	Q89906 beet yellow
33	33	66.0	3074	12	Q08534	Q08534 sugar beet
34	33	66.0	3090	12	Q9Q703	Q9q703 beet yellow
35	32	64.0	180	2	Q99VU2	Q99vu2 staphylococ
36	32	64.0	188	2	P95667	P95667 clostridium
37	32	64.0	230	5	O45766	O45766 caenorhabdi
38	32	64.0	256	12	O66182	O66182 red clover
39	32	64.0	280	2	P71102	P71102 curcubacter
40	32	64.0	305	3	P87230	P87230 schizosacch
41	32	64.0	382	9	Q38578	Q38578 bacterioph
42	32	64.0	435	13	Q9YH25	Q9yh25 brachydanio
43	32	64.0	490	13	Q9YH24	Q9yh24 brachydanio
44	32	64.0	524	11	O9EQY1	O9eqy1 rattus norv
45	32	64.0	526	10	Q9LG01	Q9lg01 arabidopsis

ALIGNMENTS

RESULT 1

Q9FHA3 ID Q9FHA3 PRELIMINARY: PRT; 1492 AA.
AC Q9FHA3
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DNA POLYMERASE ALPHA 1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3002;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:31-63(2000).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N
CC PYROPHOSPHATE + DNA(N).
CC -!- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL; AB020742; BAB10944.1.; -!
DR InterPro; IPR002064; DNA_pol_B.
DR Pfam; PF00136; DNA_pol_B; 1.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLBc; 1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase.
SQ SEQUENCE 1492 AA; 167143 MW; 27E427EED956F83 CRC64;

Query Match 76.0%; Score 38; DB 10; Length 1492;
Best Local Similarity 87.5%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 YSRKDMPL 8
Db 476 YSRKDRPL 483

RESULT 2
ID Q15725 PRELIMINARY; PRT: 521 AA.
AC Q15725;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NERF-1A-(ETS FAMILY TRANSCRIPTION FACTOR ELF2B).
GN NERF-1A,B OR ELF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11;
RP SEQUENCE FROM N.A.
RX MDLINE=96347578; PubMed=8756667;
RA Oettgen P., Akbarali Y., Boltax J., Best J., Kunsch C.,
RA Libermann T.A.;
RT "Characterization of NERF, a novel transcription factor related to the
RT Ets factor ELF-1."
RL Mol. Cell. Biol. 16:5091-5106(1996).
RN 12;
RP SEQUENCE FROM N.A.
RX MDLINE=97154704; PubMed=9001422;
RA Wilkinson D.A., Neale G.A., Mao S., Naeve C.W., Goorha R.M.;
RA Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RT "Elf-2, a rhombotin-2 binding ets transcription factor: discovery and
RT potential role in T cell leukemia."
RL Leukemia 11:86-96(1997).
RN 13;
RP SEQUENCE FROM N.A.
RA Wilkinson D.A., Neale G.A., Mao S., Fernandes E.R., Davenport J.W.,
RA Naeve C.W., Goorha R.M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U43189; AAB37760.1; -.
DR EMBL; AF256223; AAF67196.1; -.
DR HSSP; P14921; 2STT.
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF_ETs.
DR Pfam; PF00178; Ets; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS00346; ETS_DOMAIN_3; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
SQ SEQUENCE 521 AA; 56133 MW; A54F8DAA008C164D CRC64;

Query Match 72.0%; Score 36; DB 4; Length 521;
Best Local Similarity 85.7%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 YSRKDMPL 7
Db 215 YSRKDMPL 221

RESULT 3
ID Q9JHC8 PRELIMINARY; PRT: 521 AA.
AC Q9JHC8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ETS FAMILY TRANSCRIPTION FACTOR ELF2B1.
GN ELF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

Query Match 72.0%; Score 36; DB 4; Length 533;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 YSRKDMPL 7
Db 215 YSRKDMPL 221

RESULT 4
ID Q15724 PRELIMINARY; PRT: 533 AA.
AC Q15724;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE NERF-1B.
GN NERF-1A,B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11;
RP SEQUENCE FROM N.A.
RX MDLINE=96347578; PubMed=8756667;
RA Oettgen P., Akbarali Y., Boltax J., Best J., Kunsch C.,
RA Libermann T.A.;
RT "Characterization of NERF, a novel transcription factor related to the
RT Ets factor ELF-1."
RL Mol. Cell. Biol. 16:5091-5106(1996).
DR EMBL; U43189; AAB37761.1; -.
DR HSSP; P14921; 2STT.
DR InterPro; IPR000418; Ets.
DR InterPro; IPR002341; HSF_ETs.
DR Pfam; PF00178; Ets; 1.
DR PRINTS; PR00454; ETSDOMAIN.
DR SMART; SM00413; ETS; 1.
DR PROSITE; PS00345; ETS_DOMAIN_1; 1.
DR PROSITE; PS00346; ETS_DOMAIN_2; 1.
DR PROSITE; PS50061; ETS_DOMAIN_3; 1.
SQ SEQUENCE 533 AA; 57388 MW; DA45A5D12ECFD7E6 CRC64;

Query Match 72.0%; Score 36; DB 4; Length 533;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 1 YSRKDMPL 7

```

```

DB 227 YQFKDMP 233
| | | | |
RESULT 5
Q9JHC7 PRELIMINARY; PRT: 533 AA.
AC Q9JHC7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ETS FAMILY TRANSCRIPTION FACTOR ELF2B2.
GN ELF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97154704; PubMed=9001422;
RA Wilkinson D.A., Neale G.A., Mao S., Fernandes E.R., Davenport J.W.,
RA Naevae C.W., Goorha R.M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: U43188; AAB37759.1;
DR EMBL: AF256222; AAF67195.1;
DR HSP: P14921; 25TT.
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETs.
DR Pfam: PF00178; Ets; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS00061; ETS_DOMAIN_3; 1.
SQ SEQUENCE 581 AA; 62711 MW; 5D5E3169FF486A90 CRC64;

Query Match 72.0%; Score 36; DB 4; Length 581;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 1 YSFKDMP 7
| | | | |
Db 275 YQFKDMP 281

RESULT 7
Q9JHD0 PRELIMINARY; PRT: 581 AA.
ID Q9JHD0
AC Q9JHD0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ETS FAMILY TRANSCRIPTION FACTOR ELF2AL.
GN ELF2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97154704; PubMed=9001422;
RA Wilkinson D.A., Neale G.A., Mao S., Naevae C.W., Goorha R.M.;
RA "Elf-2, a rhombotin-2 binding ets transcription factor: discovery and
RT potential role in T cell leukemia.";
RL Leukemia 11:86-96(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX Wilkinson D.A., Neale G.A., Mao S., Fernandes E.R., Davenport J.W.,
RA Naevae C.W., Goorha R.M.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF256216; AAF67191.1;
DR InterPro: IPR000418; Ets.
DR InterPro: IPR002341; HSF_ETs.
DR Pfam: PF00178; Ets; 1.
DR PRINTS: PR00454; ETSDOMAIN.
DR SMART: SM00413; ETS; 1.
DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
DR PROSITE: PS00061; ETS_DOMAIN_3; 1.
SQ SEQUENCE 581 AA; 61933 MW; A39A8AB0170504ED CRC64;

Query Match 72.0%; Score 36; DB 11; Length 581;
Best Local Similarity 85.7%; Pred. No. 30;

DB 227 YQFKDMP 233
| | | | |
RESULT 6
Q15723 PRELIMINARY; PRT: 581 AA.
ID Q15723
AC Q15723
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ETS TRANSCRIPTION FACTOR (ETS FAMILY TRANSCRIPTION FACTOR ELF2A).
GN NERF-2 OR ELF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=SPLEEN, FETAL LIVER, AND FETAL BRAIN;
RX MEDLINE=96347578; PubMed=8756667;
RA Oettingen P., Akbarali Y., Bollax J., Best J., Kunsch C.,
RA Libermann T.A.;
RT "Characterization of NERF, a novel transcription factor related to the
RT ets factor ELF-1.";
RL Mol. Cell. Biol. 16:5091-5106(1996).
RN [2]

```


Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMPL 7
 DB 275 YQFKDMP 281
 RESULT 8
 ID Q9JHC9 PRELIMINARY; PRT; 593 AA.
 AC Q9JHC9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ETS FAMILY TRANSCRIPTION FACTOR ELF2A2.
 CN 2610036A20RIK OR ELF2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97154704; PubMed=9001422;
 RA Wilkinson D.A., Neale G.A., Mao S., Naeye C.W., Goorha R.M.;
 RA "Elf-2, a rhombotin-2 binding ets transcription factor: discovery and
 RT potential role in T cell leukemia.";
 RL Leukemia 11:86-96(1997).
 [2]
 RP SEQUENCE FROM N.A.
 RA Wilkinson D.A., Neale G.A.M., Mao S., Fernandes E.R., Davenport J.W.,
 RA Naeye C.W., Goorha R.M.;
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF256217; AAF67152.1;
 DR MGD: MGI:1916507; 2610036A20RIK.
 DR InterPro: IPR000418; Ets.
 DR Pfam: PF00178; Ets; 1.
 DR PRINTS: PR00454; ETS_DOMAIN.
 DR SMART: SM00413; ETS; 1.
 DR PROSITE: PS00345; ETS_DOMAIN_1; 1.
 DR PROSITE: PS00346; ETS_DOMAIN_2; 1.
 DR PROSITE: PS50061; ETS_DOMAIN_3; 1.
 SQ SEQUENCE 593 AA: 63202 MW: A6E67486141AB903 CRC64;

Query Match 72.0%; Score 36; DB 11; Length 593;
 Best Local Similarity 85.7%; Pred. No. 30;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMPL 7
 DB 287 YQFKDMP 293
 RESULT 9
 ID Q9XJB1 PRELIMINARY; PRT; 914 AA.
 AC Q9XJB1;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE PUTATIVE TAIL-HOST SPECIFICITY PROTEIN.
 OS Streptococcus thermophilus bacteriophage Drl.
 OC viruses.
 OX NCBI_TaxID=90410;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DTL;
 RX MEDLINE=99160757; PubMed=10049822;
 RA Tremblay D.M., Moineau S.;
 RA "Complete genomic sequence of the lytic bacteriophage Drl of
 RT Streptococcus thermophilus.";
 RL Virology 255:63-76(1999).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DTL;
 RA Tremblay D.M., Moineau S.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF085222; AAD21894.1;
 SQ SEQUENCE 914 AA: 103024 MW: 0C747BC05A3932C8 CRC64;

Query Match 72.0%; Score 36; DB 9; Length 914;
 Best Local Similarity 75.0%; Pred. No. 49;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMPL 8
 DB 259 YSFKDSMP 266
 RESULT 10
 ID O64284 PRELIMINARY; PRT; 1276 AA.
 AC O64284;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 142.6 KDA PROTEIN.
 OS Streptococcus thermophilus bacteriophage Sfi21.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
 OX NCBI_TaxID=64186;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98160788; PubMed=9499809;
 RA Desiere F., Lucchini S., Brussow H.;
 RA "Evolution of Streptococcus thermophilus bacteriophage genomes by
 RT modular exchanges followed by point mutations and small deletions and
 RT insertions.";
 RL Virology 241:345-356(1998).
 DR EMBL: AF115103; AAC39283.1;
 DR InterPro: IPR000087; Collagen.
 KW Hypothetical protein.
 SQ SEQUENCE 1276 AA: 142647 MW: 7F23F872B682AABD CRC64;

Query Match 72.0%; Score 36; DB 9; Length 1276;
 Best Local Similarity 75.0%; Pred. No. 72;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMPL 8
 DB 259 YSFKDSMP 266
 RESULT 11
 ID O64295 PRELIMINARY; PRT; 1291 AA.
 AC O64295;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 144.3 KDA PROTEIN.
 OS Streptococcus thermophilus bacteriophage Sfi19.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.
 OX NCBI_TaxID=72638;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98160788; PubMed=9499809;
 RA Desiere F., Lucchini S., Brussow H.;
 RA "Evolution of Streptococcus thermophilus bacteriophage genomes by
 RT modular exchanges followed by point mutations and small deletions and
 RT insertions.";
 RL Virology 241:345-356(1998).
 DR EMBL: AF115102; AAC39297.1;
 DR InterPro: IPR000087; Collagen.

```

DR Pfam: PF01391; Collagen; 1.
KW Hypothetical protein.
SQ SEQUENCE 1291 AA; 144338 MW; 571F36855F97B34F CRC64;

Query Match 72.0%; Score 36; DB 9; Length 1291;
Best Local Similarity 75.0%; Pred. No. 73;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMPX 8
Db 259 YSFKDSPM 266

RESULT 12
Q92611 ID Q92611 PRELIMINARY: PRT: 175 AA.
AC Q92611
DT 01-NOV-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 20.5 KDA PROTEIN.
OS Borrelia turicatae.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=142;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oz1;
RA Penington P.M., Cadavid D., Bunikis K., Norris S., Barbour A.G.;
RT "Interplasmidic arm duplications determine virulence phenotype of the bacterium Borrelia turicatae.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Oz1;
RA Penington P.M., Cadavid D., Bunikis J., Norris S., Barbour A.G.;
RT "Extensive interplasmidic duplications change the virulence phenotype of the relapsing fever agent Borrelia turicatae.";
RL Mol. Microbiol. 43:1120-1132(1999).
DR EMBL: AF049852; AADI5784.1;
DR EMBL: AF129434; AAF34130.1;
KW Hypothetical protein.
SQ SEQUENCE 175 AA; 20499 MW; AB038558C63F1F79 CRC64;

Query Match 70.0%; Score 35; DB 2; Length 175;
Best Local Similarity 60.0%; Pred. No. 12;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSFKDMPX 10
Db 3 YSFKDLKLT 12

RESULT 13
Q9XA26 ID Q9XA26 PRELIMINARY: PRT: 473 AA.
AC Q9XA26;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE ENDONUCLEASE.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58;
RA Claus H., Friedrich A., Frosch M., Vogel U.;
RT "Restriction barrier for transformation in Neisseria meningitidis.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ132413; CAB51246.1;
KW Endonuclease.

us-09-446-109a-3.rspt

SO SEQUENCE 473 AA; 55108 MW; 29E93256C94B9995 CRC64;

Query Match 70.0%; Score 35; DB 2; Length 473;
Best Local Similarity 75.0%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMPX 8
Db 267 YSFSEMPX 274

RESULT 14
Q9K091 ID Q9K091 PRELIMINARY: PRT: 489 AA.
AC Q9K091;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE TYPE II RESTRICTION ENZYME HGA1.
GN NMB0726.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Hatt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizzi M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.";
RL Science 287:1809-1815(2000).
DR EMBL: AE002427; AAF41139.1;
DR TIGR: NMB0726;
KW Complete proteome.
SQ SEQUENCE 489 AA; 56756 MW; 3DD90B942BCEFF106 CRC64;

Query Match 70.0%; Score 35; DB 2; Length 489;
Best Local Similarity 75.0%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMPX 8
Db 267 YSFSEMPX 274

RESULT 15
Q9DAG5 ID Q9DAG5 PRELIMINARY: PRT: 247 AA.
AC Q9DAG5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 1700011E24RIK PROTEIN.
GN 1700011E24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eumetazoa; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Ishii Y., Arakawa T., Hara A., Fukuiishi Y., Konno H., Adachi J., Fukuda S.,
```

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Gato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Welz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
DR EMBL: AK005861; BAB24281.1: -;
DR MGD: MGI:1922717; 1700011E24Rik.
SQ SEQUENCE 247 AA; 27631 MW; 5A9BBD78EE7BD9FA CRC64;

Query Match 68.0%; Score 34; DB 11; Length 247;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMP 7
| | | | |
Db 116 YSFKDKP 122

Search completed: February 27, 2002, 11:50:03
Job time: 982 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:02 ; Search time 132.19 Seconds
(without alignments)
1.702 Million cell updates/sec

Title: US-09-446-109A-3
Perfect score: 50
Sequence: 1 YSFKDMPLXR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	49	98.0	10	1	US-08-299-285-7
2	49	98.0	10	1	US-08-299-285-10
3	49	98.0	10	2	US-08-985-126-7
4	49	98.0	10	2	US-08-985-126-10
5	49	98.0	10	5	PCT-US95-11126-7
6	49	98.0	10	5	PCT-US95-11126-10
7	48	96.0	10	1	US-08-299-285-5
8	48	96.0	10	1	US-08-299-285-13
9	48	96.0	10	2	US-08-985-126-5
10	48	96.0	10	2	US-08-985-126-13
11	48	96.0	10	5	PCT-US95-11126-5
12	48	96.0	10	5	PCT-US95-11126-13
13	47	94.0	10	1	US-08-299-285-14
14	47	94.0	10	2	US-08-985-126-14
15	47	94.0	10	5	PCT-US95-11126-14
16	44	88.0	10	1	US-08-299-285-12
17	44	88.0	10	2	US-08-985-126-12
18	44	88.0	10	5	PCT-US95-11126-12
19	43	86.0	10	1	US-08-299-285-11
20	43	86.0	10	1	US-08-299-285-19
21	43	86.0	10	2	US-08-985-126-11
22	43	86.0	10	2	US-08-985-126-19
23	43	86.0	10	5	PCT-US95-11126-11
24	43	86.0	10	5	PCT-US95-11126-19
25	42	84.0	10	1	US-08-299-285-15
26	42	84.0	10	1	US-08-299-285-16
27	42	84.0	10	2	US-08-985-126-15

28	42	84.0	10	2	US-08-985-126-16	Sequence 16, Appl
29	42	84.0	10	5	PCT-US95-11126-15	Sequence 15, Appl
30	42	84.0	10	5	PCT-US95-11126-16	Sequence 16, Appl
31	41	82.0	10	1	US-08-299-285-6	Sequence 6, Appl
32	41	82.0	10	1	US-08-299-285-8	Sequence 8, Appl
33	41	82.0	10	1	US-08-299-285-17	Sequence 17, Appl
34	41	82.0	10	1	US-08-299-285-18	Sequence 18, Appl
35	41	82.0	10	1	US-08-299-285-23	Sequence 23, Appl
36	41	82.0	10	1	US-08-299-285-26	Sequence 26, Appl
37	41	82.0	10	1	US-08-299-285-27	Sequence 27, Appl
38	41	82.0	10	1	US-08-299-285-28	Sequence 28, Appl
39	41	82.0	10	1	US-08-299-285-29	Sequence 29, Appl
40	41	82.0	10	1	US-08-299-285-30	Sequence 30, Appl
41	41	82.0	10	1	US-08-299-285-31	Sequence 31, Appl
42	41	82.0	10	1	US-08-299-285-32	Sequence 32, Appl
43	41	82.0	10	1	US-08-299-285-33	Sequence 33, Appl
44	41	82.0	10	1	US-08-299-285-34	Sequence 34, Appl
45	41	82.0	10	5	PCT-US95-11126-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-08-299-285-7
; Sequence 7, Application US/08299285
; Patent No. 5696230
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
US-08-299-285-7

Query Match 98.0%; Score 49; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00023;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMPLEX 10
| | | | | | | | | |
Db 1 YSFKDMPLEX 10

RESULT 2

US-08-299-285-10
: Sequence 10, Application US/08299285
: Patent No. 5696230
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,285
: FILING DATE: 31-AUG-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: "D-Alanine at position 9"

US-08-299-285-10

Query Match 98.0%; Score 49; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00023;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMPLEX 10
| | | | | | | | | |
Db 1 YSFKDMPLEX 10

RESULT 3

US-08-985-126-7
: Sequence 7, Application US/08985126
: Patent No. 5942599
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US

APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: NO. 5942599 Relevant
TOPOLOGY: NO. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-985-126-7

Query Match 98.0%; Score 49; DB 2; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00023;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMPLEX 10
| | | | | | | | | |
Db 1 YSFKDMPLEX 10

RESULT 4

US-08-985-126-10
: Sequence 10, Application US/08985126
: Patent No. 5942599
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US

```
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/985,126
; FILING DATE:
; APPLICATION NUMBER: US/08/985,126
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: No. 5942599 Relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: "D-Alanine at position 9"
; US-08-985-126-10

Query Match 98.0%; Score 49; DB 2; Length 10:
Best Local Similarity 90.0%; Pred. No. 0.00023;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0:

QY 1 YSFKDMPLXR 10
Db 1 YSFKDMPLAR 10

RESULT 5
PCT-US95-11126-7
; Sequence 7, Application PC/TUS9511126
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
```

```
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/985,126
; FILING DATE:
; APPLICATION NUMBER: US/08/985,126
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; PCT-US95-11126-7

Query Match 98.0%; Score 49; DB 5; Length 10:
Best Local Similarity 90.0%; Pred. No. 0.00023;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0:

QY 1 YSFKDMPLXR 10
Db 1 YSFKDMPLAR 10

RESULT 6
PCT-US95-11126-10
; Sequence 10, Application PC/TUS9511126
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
```

STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /product= "D-Alanine at position 9"

PCT-US95-11128-10

Query Match: 98.0%; Score 49; DB 5; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00023;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 1 YSFKDMPLEXR 10
Db 1 YSFKDMPLEXR 10

RESULT 7

US-08-299-285-5
; Sequence 5, Application US/08299285
; Patent No. 5696230
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
US-08-299-285-5

Query Match 96.0%; Score 48; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00037;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 1 YSFKDMPLEXR 10
Db 1 YSFKDMPLEXR 10

RESULT 8

US-08-299-285-13
; Sequence 13, Application US/08299285
; Patent No. 5696230
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: "D-Phenylalanine at position 9"
US-08-299-285-13

Query Match 96.0%; Score 48; DB 1; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00037;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 1 YSFKDMPLEXR 10
Db 1 YSFKDMPLEXR 10

RESULT 9

US-08-985-126-5
; Sequence 5, Application US/08985126
; Patent No. 5942599
; GENERAL INFORMATION:

APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: NO. 5942599 Relevant
TOPOLOGY: NO. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-985-126-5

Query Match 96.0%; Score 48; DB 2; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00037;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMPLXR 10
Db 1 YSFKDMPLGR 10

RESULT 10
US-08-985-126-13
Sequence 13, Application US/08985126
Patent No. 5942599
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US

ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: NO. 5942599 Relevant
TOPOLOGY: NO. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: "D-Phenylalanine at
position 9"
US-08-985-126-13

Query Match 96.0%; Score 48; DB 2; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00037;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 YSFKDMPLXR 10
Db 1 YSFKDMPLFR 10

RESULT 11
PCT-US95-11126-5
Sequence 5, Application PC/TUS9511126
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11126
FILING DATE:


```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; PCT-US95-11126-5

Query Match 96.0%; Score 48; DB 5; Length 10;
Best Local Similarity 90.0%; Pred. No. 0.00037;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSFKDMPLEXR 10
Db 1 YSFKDMPLEXR 10

RESULT 12
PCT-US95-11126-13
; Sequence 13, Application PC/TUS9511126
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; NAME/KEY: Modified-site

PCT-US95-11126-13
; Sequence 14, Application US/08299285
; Patent No. 5696230
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; NAME/KEY: Modified-site

PCT-US95-11126-13
; Sequence 14, Application US/08299285
; Patent No. 5696230
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; NAME/KEY: Modified-site
```

LOCATION: 9
OTHER INFORMATION: "D-proline at position 9"
US-08-299-285-14

Query Match 94.0% Score 47; DB 1; Length 10;
Best Local Similarity 90.0% Pred. No. 0.0006;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSEKDMPLXR 10
IIIIIIII I
Db 1 YSEKDMPLR 10

RESULT 14
US-08-985-126-14
Sequence 14, Application US/08985126
Patent No. 5942599
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: "D-proline at position 9"
US-08-985-126-14

Query Match 94.0% Score 47; DB 2; Length 10;
Best Local Similarity 90.0% Pred. No. 0.0006;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSEKDMPLXR 10
IIIIIIII I
Db 1 YSEKDMPLR 10

RESULT 15
PCT-US95-11126-14
Sequence 14, Application PC/TUS9511126
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: "D-proline at position 9"
PCT-US95-11126-14

Query Match 94.0% Score 47; DB 5; Length 10;
Best Local Similarity 90.0% Pred. No. 0.0006;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSEKDMPLXR 10
IIIIIIII I
Db 1 YSEKDMPLR 10

Search completed: February 27, 2002, 11:36:02
Job time: 141 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:13 ; Search time 303.5 Seconds
(without alignments)
2.441 Million cell updates/sec

Title: us-09-446-109A-4
Perfect score: 51
Sequence: 1 YSEKPLXR 10

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101:*

1: /SID58/qcgdata/geneseq/geneseq/AA1980.DAT:*

2: /SID58/qcgdata/geneseq/geneseq/AA1981.DAT:*

3: /SID58/qcgdata/geneseq/geneseq/AA1982.DAT:*

4: /SID58/qcgdata/geneseq/geneseq/AA1983.DAT:*

5: /SID58/qcgdata/geneseq/geneseq/AA1984.DAT:*

6: /SID58/qcgdata/geneseq/geneseq/AA1985.DAT:*

7: /SID58/qcgdata/geneseq/geneseq/AA1986.DAT:*

8: /SID58/qcgdata/geneseq/geneseq/AA1987.DAT:*

9: /SID58/qcgdata/geneseq/geneseq/AA1988.DAT:*

10: /SID58/qcgdata/geneseq/geneseq/AA1989.DAT:*

11: /SID58/qcgdata/geneseq/geneseq/AA1990.DAT:*

12: /SID58/qcgdata/geneseq/geneseq/AA1991.DAT:*

13: /SID58/qcgdata/geneseq/geneseq/AA1992.DAT:*

14: /SID58/qcgdata/geneseq/geneseq/AA1993.DAT:*

15: /SID58/qcgdata/geneseq/geneseq/AA1994.DAT:*

16: /SID58/qcgdata/geneseq/geneseq/AA1995.DAT:*

17: /SID58/qcgdata/geneseq/geneseq/AA1996.DAT:*

18: /SID58/qcgdata/geneseq/geneseq/AA1997.DAT:*

19: /SID58/qcgdata/geneseq/geneseq/AA1998.DAT:*

20: /SID58/qcgdata/geneseq/geneseq/AA1999.DAT:*

21: /SID58/qcgdata/geneseq/geneseq/AA2000.DAT:*

22: /SID58/qcgdata/geneseq/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	98.0	10	AAW27250	Molecular adjuvant
2	50	98.0	12	AAW27254	Molecular adjuvant
3	50	98.0	18	AAW27256	Molecular adjuvant
4	50	98.0	19	AAW27251	Molecular adjuvant
5	42	82.4	10	AAW27251	C5a anaphylatoxin
6	42	82.4	10	AAW27251	C5a anaphylatoxin
7	42	82.4	10	AAW27251	C5a anaphylatoxin
8	41	80.4	10	AAW27251	C5a anaphylatoxin
9	41	80.4	10	AAW27251	C5a anaphylatoxin
10	40	78.4	10	AAW27251	C5a anaphylatoxin
11	40	78.4	10	AAW27251	C5a anaphylatoxin

12	37	72.5	10	17	AAW27250	C5a anaphylatoxin
13	36	70.6	10	17	AAW27250	C5a anaphylatoxin
14	36	70.6	10	17	AAW27250	C5a anaphylatoxin
15	35	68.6	10	17	AAW27250	C5a anaphylatoxin
16	35	68.6	10	17	AAW27250	C5a anaphylatoxin
17	35	68.6	263	21	AAW27250	Arabidopsis thaliana
18	35	68.6	280	21	AAW27250	Arabidopsis thaliana
19	35	68.6	285	21	AAW27250	Arabidopsis thaliana
20	35	68.6	504	21	AAW27250	Arabidopsis thaliana
21	35	68.6	504	21	AAW27250	Arabidopsis thaliana
22	35	68.6	509	21	AAW27250	Caenorhabditis elegans
23	35	68.6	510	21	AAW27250	Caenorhabditis elegans
24	35	68.6	515	21	AAW27250	Arabidopsis thaliana
25	35	68.6	515	21	AAW27250	Arabidopsis thaliana
26	35	68.6	774	22	AAW27250	Human protein sequence
27	34	66.7	10	17	AAW27250	C5a anaphylatoxin
28	34	66.7	10	17	AAW27250	C5a anaphylatoxin
29	34	66.7	10	17	AAW27250	C5a anaphylatoxin
30	34	66.7	10	17	AAW27250	C5a anaphylatoxin
31	34	66.7	10	17	AAW27250	C5a anaphylatoxin
32	34	66.7	10	17	AAW27250	C5a anaphylatoxin
33	34	66.7	10	17	AAW27250	C5a analogue C5a(6)
34	34	66.7	10	17	AAW27250	C5a analogue C5a(6)
35	34	66.7	10	17	AAW27250	C5a analogue C5a(6)
36	34	66.7	10	17	AAW27250	C5a analogue C5a(6)
37	34	66.7	10	17	AAW27250	C5a analogue C5a(6)
38	34	66.7	10	17	AAW27250	C5a anaphylatoxin
39	34	66.7	10	17	AAW27250	C5a anaphylatoxin
40	34	66.7	10	17	AAW27250	C5a anaphylatoxin
41	34	66.7	29	18	AAW27250	Anti-kappa(33-52)
42	34	66.7	186	21	AAW27250	Arabidopsis thaliana
43	34	66.7	253	21	AAW27250	Arabidopsis thaliana
44	34	66.7	253	21	AAW27250	Arabidopsis thaliana
45	34	66.7	271	21	AAW27250	Arabidopsis thaliana

ALIGNMENTS

RESULT 1

AAW27250

ID AAW27250 standard; peptide; 10 AA.

XX AC AAW27250;

XX DT 23-DEC-1997 (first entry)

XX DE Molecular adjuvant targeting ligand.

XX KW Molecular adjuvant; immune response; immunogen; binding affinity; antigen presenting cell; APC; viral pathogen; anti-tumour response; antibody; immunodiagnostic; immunotherapeutic; hist.

XX OS Synthetic.

XX FH Key

XX FT Misc-difference 9

XX FT /note= "D-form residue"

XX PW W09714426-A1.

XX PD 24-APR-1997.

XX PF 18-OCT-1996; 96WO-US16825.

XX PR 20-OCT-1995; 95US-0005727.

XX PA (UYNE-) UNIV NEBRASKA.

XX PI Hollingsworth MA, Sanderson SD, Tempero RA;

XX DR WPI; 1997-244854/22.

PT Molecular adjuvants for enhancing immune responses - comprise an
 PT immunogen linked to a ligand having binding affinity for a
 PT determinant of an antigen presenting cell

XX Claim 5; Page 51; 61pp; English.

XX A novel molecular adjuvant has been developed for enhancing an immune
 CC response to an immunogen. The molecular adjuvant comprises a targeting
 CC ligand, having binding affinity for a characteristic determinant of an
 CC antigen presenting cell (APC), the targeting ligand being functionally
 CC linked to the immunogen, and so binding of the molecular adjuvant to
 CC the APC determinant activates the APC, effecting delivery of the
 CC immunogen to an antigen presenting pathway of the APC. The present
 CC sequence represents a specifically claimed targeting ligand, which is
 CC a C5a C-terminal decapeptide agonist. The molecular adjuvant can be used
 CC to produce an immune response for protecting against viral and other
 CC pathogens or to produce anti-tumour responses. It can also be used for
 CC the production of antibodies for use as immunodiagnostic and
 CC immunotherapeutic agents. The molecular adjuvant is readily taken up and
 CC processed by APCs to provide APC-mediated immune responses. It can
 CC provide for the production of antibodies against a weakly-antigenic or
 CC non-antigenic substances.

XX Sequence 10 AA:

Query Match 98.0%; Score 50; DB 18; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.0023;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSRKPMPLXR 10
 Db 1 ysfkpmplar 10
 |||||

RESULT 2

AAW27254
 ID AAW27254 standard; peptide; 12 AA.

AC AAW27254;

DT 23-DEC-1997 (first entry)

DE Molecular adjuvant formula.

XX Molecular adjuvant; immune response; immunogen; binding affinity;
 KW antigen presenting cell; APC; viral pathogen; anti tumour response;
 KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note= "Serum amyloid A is attached to the N-terminal"

FT Modified-site 2 /label= OTHER

FT /note= "Aminohexanoic acid which is a linear
 aliphatic spacer moiety"

FT Misc-difference 11 /note= "D-form residue"

XX W09714426-A1.

PD 24-APR-1997.

XX 18-OCT-1996; 96WO-US16825.

XX 20-OCT-1995; 95US-0005727.

XX (UYNE-) UNIV NEBRASKA.

PA Hollingsworth MA, Sanderson SD, Tempero RA;

XX

WPI; 1997-244854/22.

XX Molecular adjuvants for enhancing immune responses - comprise an
 PT immunogen linked to a ligand having binding affinity for a
 PT determinant of an antigen presenting cell

XX Claim 15; Page 53; 61pp; English.

XX A novel molecular adjuvant has been developed for enhancing an immune
 CC response to an immunogen. The molecular adjuvant comprises a targeting
 CC ligand, having binding affinity for a characteristic determinant of an
 CC antigen presenting cell (APC), the targeting ligand being functionally
 CC linked to the immunogen, and so binding of the molecular adjuvant to
 CC the APC determinant activates the APC, effecting delivery of the
 CC immunogen to an antigen presenting pathway of the APC. The present
 CC sequence represents a specifically claimed formula for a molecular
 CC adjuvant, where the immunogen is serum amyloid A. The molecular adjuvant
 CC can be used to produce an immune response for protecting against viral
 CC and other pathogens or to produce anti-tumour responses. It can also be
 CC used for the production of antibodies for use as immunodiagnostic and
 CC immunotherapeutic agents. The molecular adjuvant is readily taken up and
 CC processed by APCs to provide APC-mediated immune responses. It can
 CC provide for the production of antibodies against a weakly-antigenic or
 CC non-antigenic substances.

XX Sequence 12 AA:

Query Match 98.0%; Score 50; DB 18; Length 12;
 Best Local Similarity 90.0%; Pred. No. 0.0027;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSRKPMPLXR 10
 Db 3 ysfkpmplar 12
 |||||

RESULT 3

AAW27256
 ID AAW27256 standard; peptide; 18 AA.

AC AAW27256;

DT 23-DEC-1997 (first entry)

DE Molecular adjuvant targeting ligand and immunogen.

XX Molecular adjuvant; immune response; immunogen; binding affinity;
 KW antigen presenting cell; APC; viral pathogen; anti tumour response;
 KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 9 /note= "D-form residue"

XX W09714426-A1.

XX 24-APR-1997.

XX 18-OCT-1996; 96WO-US16825.

XX 20-OCT-1995; 95US-0005727.

XX (UYNE-) UNIV NEBRASKA.

XX Hollingsworth MA, Sanderson SD, Tempero RA;

XX WPI; 1997-244854/22.

XX Molecular adjuvants for enhancing immune responses - comprise an
 PT immunogen linked to a ligand having binding affinity for a

PT determinant of an antigen presenting cell

XX Example 1: Page 31: 61pp; English.

XX A novel molecular adjuvant has been developed for enhancing an immune
CC response to an immunogen. The molecular adjuvant comprises a targeting
CC ligand, having binding affinity for a characteristic determinant of an
CC antigen presenting cell (APC), the targeting ligand being functionally
CC linked to the immunogen, and so binding of the molecular adjuvant to
CC the APC determinant activates the APC, effecting delivery of the
CC immunogen to an antigen presenting pathway of the APC. The present
CC sequence represents a targeting ligand and immunogen, in which the
CC junctional epitope of human mucin-1 is positioned toward the
CC carboxyl terminus and the C5a analogue is positioned toward the amino
CC terminus. The molecular adjuvant can be used to produce an immune
CC response for protecting against viral and other pathogens or to
CC produce anti-tumour responses. It can also be used for the production
CC of antibodies for use as immunodiagnostic and immunotherapeutic
CC agents. The molecular adjuvant is readily taken up and processed
CC by APCs to provide APC-mediated immune responses. It can provide
CC for the production of antibodies against a weakly-antigenic or
CC non-antigenic substances.

XX Sequence 18 AA;

Query Match 98.0%; Score 50; DB 18; Length 18;
Best Local Similarity 90.0%; Pred. No. 0.0041;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSKPKMPLXR 10
Db | | | | | | | | | |
1 yskkpmplar 10

RESULT 4

AAW27251
ID AAW27251 standard; peptide: 19 AA.

XX AC AAW27251;

XX 23-DEC-1997 (first entry)

XX Molecular adjuvant targeting ligand and immunogen.

XX Molecular adjuvant; immune response; immunogen: binding affinity;
KW antigen presenting cell; APC; viral pathogen; anti tumour response;
KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX Synthetic.

XX Key Location/Qualifiers
FH Misc-difference 18 /note= "D-form residue"

XX W09714426-AL.

XX 24-APR-1997.

XX 18-OCT-1996; 96WO-US16825.

XX 20-OCT-1995; 95US-0005727.

XX (UYNE-) UNIV NEBRASKA.

XX Hollingsworth MA, Sanderson SD, Tempero RA;

XX WPI; 1997-244854/22.

XX Molecular adjuvants for enhancing immune responses - comprise an
PT immunogen linked to a ligand having binding affinity for a
PT determinant of an antigen presenting cell

PS Claim 6: Page 52; 61pp; English.

XX A novel molecular adjuvant has been developed for enhancing an immune
CC response to an immunogen. The molecular adjuvant comprises a targeting
CC ligand, having binding affinity for a characteristic determinant of an
CC antigen presenting cell (APC), the targeting ligand being functionally
CC linked to the immunogen, and so binding of the molecular adjuvant to
CC the APC determinant activates the APC, effecting delivery of the
CC immunogen to an antigen presenting pathway of the APC. The present
CC sequence represents a specifically claimed targeting ligand and
CC immunogen. The molecular adjuvant can be used to produce an immune
CC response for protecting against viral and other pathogens or to
CC produce anti-tumour responses. It can also be used for the production
CC of antibodies for use as immunodiagnostic and immunotherapeutic
CC agents. The molecular adjuvant is readily taken up and processed
CC by APCs to provide APC-mediated immune responses. It can provide
CC for the production of antibodies against a weakly-antigenic or
CC non-antigenic substances.

XX Sequence 19 AA;

Query Match 98.0%; Score 50; DB 18; Length 19;
Best Local Similarity 90.0%; Pred. No. 0.0044;
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSKPKMPLXR 10
Db | | | | | | | | | |
10 yskkpmplar 19

RESULT 5

AAW27487
ID AAR94487 standard; peptide: 10 AA.

XX AC AAR94487;

XX 17-OCT-1996 (first entry)

XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 Pro69 D-Ala73.

XX C-terminal; human: C5a anaphylatoxin; analogue: beta-turn;
KW constrained backbone conformation; proinflammatory response;
KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
KW mediated; increase; cell membrane; vascular; permeability;
KW neutrophil polarisation; neutrophil enzyme release; treatment;
KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
KW cancer; drug delivery; CNS disorder; central nervous system;
KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
KW development; non-steroidal; anti-inflammatory.

XX Homo sapiens.

XX Key Location/Qualifiers
FH Region 4...7 /note= "beta-turn"

XX Misc-difference 10 /note= "D-form residue"

XX W09606629-AL.

XX 07-MAR-1996.

XX 31-AUG-1995; 95WO-US11126.

XX 31-AUG-1994; 94US-0299285.

XX (UYNE-) UNIV NEBRASKA.

XX (UYQU) UNIV QUEENSLAND.

XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;

XX WPI; 1996-160140/16.

XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane permeability
 PT
 XX
 XX
 PS Claim 8; Page 88; 116pp; English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 XX
 SO Sequence 10 AA;

Query Match 82.4%; Score 42; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.074; 2: Indels 0: Gaps 0;
 Matches 8: Conservative 0: Mismatches

QY 1 YSKPKMPLXR 10
 |||||
 Db 1 ysfkpmqlar 10

RESULT 6
 AAR94470
 ID AAR94470 standard; peptide: 10 AA.
 AC AAR94470;
 XX
 DT 17-OCT-1996 (first entry)
 DE
 XX C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 Ala73.
 CC
 CC C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 CC constrained backbone conformation; proinflammatory response;
 CC spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 CC mediated; increase; cell membrane; vascular; permeability;
 CC neutrophil polarisation; neutrophil enzyme release; treatment;
 CC immune adjuvant; immunodeficiency; augmentation; immune therapy;
 CC cancer; drug delivery; CNS disorder; central nervous system;
 CC Alzheimer's disease; tumour; high affinity; receptor antagonist;
 CC development; non-steroidal; anti-inflammatory.
 XX
 OS Homo sapiens.
 XX
 FI Key Location/Qualifiers
 FT Region 4..7
 FT /note= "beta-turn"
 XX
 XX WO9606629-A1.
 XX
 PD 07-MAR-1996.
 XX
 XX 31-AUG-1995; 95WO-US11126.
 XX
 XX 31-AUG-1994; 94US-0299285.
 XX
 XX (UYNE-) UNIV NEBRASKA.

(UYOU) UNIV QUEENSLAND.
 KIRNARSKY L, SANDERSON DG, SHERMAN SA, TAYLOR SM;
 WPI: 1996-160140/16.
 CC C-terminal analogues of C5a anaphylatoxin - induce e.g.
 CC spasmogenesis, platelet aggregation and increases in cell membrane permeability
 CC
 XX Example 2; Page 79; 116pp; English.
 CC
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (micromolar) activities of 122 and 48, compared to 0.018 and 0.0013
 CC for C5a.
 XX
 SO Sequence 10 AA;

Query Match 82.4%; Score 42; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.074; 2: Indels 0: Gaps 0;
 Matches 8: Conservative 0: Mismatches

QY 1 YSKPKMPLXR 10
 |||||
 Db 1 ysfkpmqlar 10

RESULT 7
 AAR94473
 ID AAR94473 standard; peptide: 10 AA.
 AC AAR94473;
 XX
 DT 17-OCT-1996 (first entry)
 DE
 XX C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 Pro71 D-Ala73.
 CC
 CC C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 CC constrained backbone conformation; proinflammatory response;
 CC spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 CC mediated; increase; cell membrane; vascular; permeability;
 CC neutrophil polarisation; neutrophil enzyme release; treatment;
 CC immune adjuvant; immunodeficiency; augmentation; immune therapy;
 CC cancer; drug delivery; CNS disorder; central nervous system;
 CC Alzheimer's disease; tumour; high affinity; receptor antagonist;
 CC development; non-steroidal; anti-inflammatory.
 XX
 OS Homo sapiens.
 XX
 FI Key Location/Qualifiers
 FT Region 4..7
 FT /note= "beta-turn"

FT Misc-difference 9 /note= "D-form residue"
 FT KW W09606629-A1.
 XX 07-MAR-1996:
 XX 31-AUG-1995: 95WO-US11126.
 XX 31-AUG-1994: 94US-0299285.
 XX (UYNE-) UNIV NEBRASKA.
 XX (UYQU) UNIV QUEENSLAND.
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI: 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX Claim 8: Page 81: 116pp: English.
 XX The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 0.35 and 2.1, compared to 0.018 and 0.0013
 CC for C5a.
 XX Sequence 10 AA:

Query Match 82.4%; Score 42; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.074;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSFKPMPLXR 10
 III III I
 Db 1 ysfkdmlpr 10

RESULT 8
 AAR94476
 ID AAR94476 standard: peptide: 10 AA.
 XX
 AC AAR94476;
 XX
 DT 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 Pro71 D-Phe73.
 XX C-terminal: human; C5a anaphylatoxin; analogue: beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;

mediated; increase: cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FT Region 4..7
 FT /note= "beta-turn"
 FT Misc-difference 9 /note= "D-form residue"
 FT W09606629-A1.
 PN 07-MAR-1996.
 PD 31-AUG-1995: 95WO-US11126.
 XX 31-AUG-1994: 94US-0299285.
 XX (UYNE-) UNIV NEBRASKA.
 XX (UYQU) UNIV QUEENSLAND.
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI: 1996-160140/16.
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX Example 2: Page 82: 116pp: English.
 XX The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 51.1 and 4.0, compared to 0.018 and 0.0013
 CC for C5a.
 XX Sequence 10 AA:

Query Match 80.4%; Score 41; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.11;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSFKPMPLXR 10
 III III I
 Db 1 ysfkdmlpr 10

RESULT 9

AAR94468
 ID AAR94468 standard; peptide; 10 AA.
 AC AAR94468;
 XX 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 Pro71.
 DE
 XX C-terminal; human; C5a anaphylatoxin; analogue; beta-turn; constrained backbone conformation; proinflammatory response; spasmogenesis; platelet aggregation; neutrophil; non-mediated; mediated; increase; cell membrane; vascular; permeability; neutrophil polarisation; neutrophil enzyme release; treatment; immune adjuvant; immunodeficiency; augmentation; immune therapy; cancer; drug delivery; CNS disorder; central nervous system; Alzheimer's disease; tumour; high affinity; receptor antagonist; development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH Region 4..7
 FT /note= "beta-turn"
 FT
 PN WO9606629-A1.
 XX
 XX 07-MAR-1996.
 XX
 XX 31-AUG-1995; 95WO-US11126.
 XX
 XX 31-AUG-1994; 94US-0299285.
 XX (UYNE-) UNIV NEBRASKA.
 PA (UYQU) UNIV QUEENSLAND.
 XX
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 PI WPI: 1996-160140/16.
 DR
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g. spasmogenesis, platelet aggregation and increases in cell membrane permeability
 PT
 PT
 XX Claim 8; Page 78; 116pp; English.
 PS
 XX The present peptide is an analogue of a C-terminal decapeptide region from human C5a anaphylatoxin, which has a constrained backbone conformation (compared to the natural peptide's flexible C-terminal region) comprising a beta-turn. Analogues with type (II or V) and type III beta-turns elicit proinflammatory responses characterised by spasmogenesis, platelet aggregation and neutrophil non-mediated increases of cell membrane (specifically vascular) permeability, or neutrophil polarisation, neutrophil enzyme release and neutrophil mediated increases in cell membrane (esp. vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of immune deficiency disorders, and for augmenting standard immune therapy for the treatment of cancer, without inflammatory side effects. They can also be used to facilitate drug delivery by increasing vascular permeability, e.g. in the treatment of CNS disorders such as Alzheimer's disease or tumours, and to develop high affinity C5a receptor antagonists useful as non-steroidal anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth muscle contraction assay, and a type III beta-turn specific neutrophil polarisation assay the peptide had respective EC50 (microM) activities of 1.13 and 5, compared to 0.018 and 0.0013 for C5a.
 CC
 XX Sequence 10 AA:
 SX

Query Match 80.4%; Score 41; DB 17
 Best Local 80.0%; Pred. No. 0.11;
 Matches 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 YSEKPMPLXR 10
 III III I
 DB 1 Ysfkdmplgr 10

RESULT 10
 AAR94486
 ID AAR94486 standard; peptide; 10 AA.
 XX
 AC AAR94486;
 XX
 DT 17-OCT-1996 (first entry)
 XX C5a anaphylatoxin analogue C5a(65-74) Tyr65 Gly69 Pro70.
 DE
 XX C-terminal; human; C5a anaphylatoxin; analogue; beta-turn; constrained backbone conformation; proinflammatory response; spasmogenesis; platelet aggregation; neutrophil; non-mediated; mediated; increase; cell membrane; vascular; permeability; neutrophil polarisation; neutrophil enzyme release; treatment; immune adjuvant; immunodeficiency; augmentation; immune therapy; cancer; drug delivery; CNS disorder; central nervous system; Alzheimer's disease; tumour; high affinity; receptor antagonist; development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH Region 4..7
 FT /note= "beta-turn"
 FT
 PN WO9606629-A1.
 XX
 XX 07-MAR-1996.
 XX
 XX 31-AUG-1995; 95WO-US11126.
 XX
 XX 31-AUG-1994; 94US-0299285.
 XX (UYNE-) UNIV NEBRASKA.
 PA (UYQU) UNIV QUEENSLAND.
 XX
 XX Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 PI WPI: 1996-160140/16.
 DR
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g. spasmogenesis, platelet aggregation and increases in cell membrane permeability
 PT
 PT
 XX Claim 8; Page 87; 116pp; English.
 PS
 XX The present peptide is an analogue of a C-terminal decapeptide region from human C5a anaphylatoxin, which has a constrained backbone conformation (compared to the natural peptide's flexible C-terminal region) comprising a beta-turn. Analogues with type (II or V) and type III beta-turns elicit proinflammatory responses characterised by spasmogenesis, platelet aggregation and neutrophil non-mediated increases of cell membrane (specifically vascular) permeability, or neutrophil polarisation, neutrophil enzyme release and neutrophil mediated increases in cell membrane (esp. vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of immune deficiency disorders, and for augmenting standard immune therapy for the treatment of cancer, without inflammatory side effects. They can also be used to facilitate drug delivery by increasing vascular permeability, e.g. in the treatment of CNS disorders such as Alzheimer's disease or tumours, and to develop high affinity C5a receptor antagonists useful as non-steroidal

CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 0.59 and 2.0, compared to 0.018 and 0.0013
 CC for C5a.
 XX Sequence 10 AA;
 SQ

Query Match 78.4%; Score 40; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.18;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSFKPMPLXR 10
 Db 1 ysfkgmplgr 10

RESULT 11
 AAR94477
 ID AAR94477 standard; peptide: 10 AA.
 XX
 AC AAR94477;
 DT 17-OCT-1996 (first entry)
 XX
 DE C5a anaphylatoxin analogue C5a(65-74) Tyr65 Pro71 D-Pro73.
 XX
 KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 4..7 /note= "beta-turn"
 FT Misc-difference 9 /note= "D-form residue"
 FT
 XX WO9606629-Al.
 PN
 PD 07-MAR-1996.
 XX
 PF 31-AUG-1995; 95WO-US11126.
 XX
 PR 31-AUG-1994; 94US-0299285.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX
 DR WPI: 1996-160140/16.
 XX
 PT C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Claim 8; Page 83; 116pp; English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses

CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 3.7 and 2.5, compared to 0.018 and 0.0013
 CC for C5a.
 XX
 SQ Sequence 10 AA;
 Query Match 78.4%; Score 40; DB 17; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.18;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSFKPMPLXR 10
 Db 1 ysfkgmplgr 10

RESULT 12
 AAR94475
 ID AAR94475 standard; peptide: 10 AA.
 XX
 AC AAR94475;
 DT 17-OCT-1996 (first entry)
 XX
 DE C5a anaphylatoxin analogue C5a(65-74) Tyr65 Pro71 Ala72 D-Ala73.
 XX
 KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Region 4..7 /note= "beta-turn"
 FT Misc-difference 9 /note= "D-form residue"
 FT
 XX WO9606629-Al.
 PN
 PD 07-MAR-1996.
 XX
 PF 31-AUG-1995; 95WO-US11126.
 XX
 PR 31-AUG-1994; 94US-0299285.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX
 DR WPI: 1996-160140/16.

XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 PT spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 XX
 PS Example 2; Page 66; 116pp; English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 17.9 and 5.6, compared to 0.018 and 0.0013
 CC for C5a.
 XX Sequence 10 AA;
 SQ

Query Match 72.5%; Score 37; DB 17; Length 10;
 Best Local Similarity 70.0%; Pred. No. 0.66;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YSKPKMPLXR 10
 IIII I I I
 Db 1 ystkdmpear 10

RESULT 13
 AAR94474
 ID AAR94474 standard; peptide; 10 AA.
 XX
 AC AAR94474;
 XX
 DT 17-OCT-1996 (first entry)
 XX
 DE C5a anaphylatoxin C-terminal analogue C5a(65-74) Tyr65 pro71 Ala72.
 XX
 KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 XX Region 4..7
 FT /note= "beta-turn"
 FT
 PN W09606629-A1.
 XX
 PD 07-MAR-1996.
 XX

PF 31-AUG-1995; 95WO-US11126.
 XX
 PR 31-AUG-1994; 94US-0299285.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 PA (UYOU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI; 1996-160140/16.
 DR
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 XX spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 PT
 XX Example 2; Page 81; 116pp; English.
 PS
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of >1000 and 215, compared to 0.018 and 0.0013
 CC for C5a.
 XX Sequence 10 AA;
 SQ

Query Match 70.6%; Score 36; DB 17; Length 10;
 Best Local Similarity 70.0%; Pred. No. 1;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 YSKPKMPLXR 10
 IIII I I I
 Db 1 ystkdmpeagr 10
 XX
 RESULT 14
 AAR94482
 ID AAR94482 standard; peptide; 10 AA.
 XX
 AC AAR94482;
 XX
 DT 17-OCT-1996 (first entry)
 XX
 DE C5a anaphylatoxin analogue C5a(65-74) Tyr65 Ala70 Pro71 D-Ala73.
 XX
 KW C-terminal; human; C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX

OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 KW Region 4..7
 FT /note= "beta-turn"
 FT Misc-difference 9
 FT /note= "D-form residue"
 XX
 PN WO9606629-A1.
 XX
 PD 07-MAR-1996.
 XX
 PF 31-AUG-1995; 95WO-US11126.
 XX
 PR 31-AUG-1994; 94US-0299285.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 PA (UYQU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI; 1996-160140/16.
 DR
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 XX spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 PT
 XX
 PS Claim 8; Page 85; 116pp; English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 0.33 and 2.3, compared to 0.018 and 0.0013
 CC for C5a.
 XX
 SQ Sequence 10 AA:

Query Match 70.6%; Score 36; DB 17; Length 10;
 Best Local Similarity 70.0%; Pred. NO. 1;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 YSFKPMPLXR 10
 |||| || |
 Db 1 ysfkdaplr 10

RESULT 15

AAR94478
 ID AAR94478 standard; peptide: 10 AA.

XX
 AC AAR94478;

XX
 DT 17-OCT-1996 (first entry)

XX

DE C5a anaphylatoxin analogue C5a(65-74).Tyr65 Ala70 Pro71.
 XX
 KW C-terminal; human: C5a anaphylatoxin; analogue; beta-turn;
 KW constrained backbone conformation; proinflammatory response;
 KW spasmogenesis; platelet aggregation; neutrophil; non-mediated;
 KW mediated; increase; cell membrane; vascular; permeability;
 KW neutrophil polarisation; neutrophil enzyme release; treatment;
 KW immune adjuvant; immunodeficiency; augmentation; immune therapy;
 KW cancer; drug delivery; CNS disorder; central nervous system;
 KW Alzheimer's disease; tumour; high affinity; receptor antagonist;
 KW development; non-steroidal; anti-inflammatory.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 KW Region 4..7
 FT /note= "beta-turn"
 FT
 XX
 PN WO9606629-A1.
 XX
 PD 07-MAR-1996.
 XX
 PF 31-AUG-1995; 95WO-US11126.
 XX
 PR 31-AUG-1994; 94US-0299285.
 XX
 PA (UYNE-) UNIV NEBRASKA.
 PA (UYQU) UNIV QUEENSLAND.
 XX
 PI Kirnarsky L, Sanderson DG, Sherman SA, Taylor SM;
 XX WPI; 1996-160140/16.
 DR
 XX C-terminal analogues of C5a anaphylatoxin - induce e.g.
 XX spasmogenesis, platelet aggregation and increases in cell membrane
 PT permeability
 PT
 XX
 PS Claim 8; Page 83; 116pp; English.
 XX
 CC The present peptide is an analogue of a C-terminal decapeptide
 CC region from human C5a anaphylatoxin, which has a constrained
 CC backbone conformation (compared to the natural peptide's flexible
 CC C-terminal region) comprising a beta-turn. Analogues with type
 CC (II or V) and type III beta-turns elicit proinflammatory responses
 CC characterised by spasmogenesis, platelet aggregation and neutrophil
 CC non-mediated increases of cell membrane (specifically vascular)
 CC permeability, or neutrophil polarisation, neutrophil enzyme
 CC release and neutrophil mediated increases in cell membrane (esp.
 CC vascular) permeability, respectively.
 CC The analogues can be used as immune adjuvants for the treatment of
 CC immune deficiency disorders, and for augmenting standard immune
 CC therapy for the treatment of cancer, without inflammatory side
 CC effects. They can also be used to facilitate drug delivery by
 CC increasing vascular permeability, e.g. in the treatment of CNS
 CC disorders such as Alzheimer's disease or tumours, and to develop
 CC high affinity C5a receptor antagonists useful as non-steroidal
 CC anti-inflammatory agents.
 CC In a type (II or V) beta-turn specific human foetal artery smooth
 CC muscle contraction assay, and a type III beta-turn specific
 CC neutrophil polarisation assay the peptide had respective EC50
 CC (microm) activities of 1.74 and 15.8, compared to 0.018 and 0.0013
 CC for C5a.
 XX
 SQ Sequence 10 AA:

Query Match 68.6%; Score 35; DB 17; Length 10;
 Best Local Similarity 70.0%; Pred. NO. 1.6;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 YSFKPMPLXR 10
 |||| || |
 Db 1 ysfkdaplr 10

Search completed: February 27, 2002, 11:41:13
Job time: 452 sec

34

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:13 ; Search time 145.23 Seconds
(without alignments)
5.245 Million cell updates/sec

Title: US-09-446-109A-4
Perfect score: >51
Sequence: 1 YSFKPMPLXR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	70.6	940	JE0291	FB19 protein - hum
2	35	68.6	495	T48372	AtB'alpha regulato
3	35	68.6	510	T42255	HNF-3/forhead pro
4	35	68.6	530	T42234	fork head-related
5	35	68.6	622	T37204	hypothetical prote
6	35	68.6	690	T23775	hypothetical prote
7	35	68.6	717	T27066	hypothetical prote
8	35	68.6	717	T27067	hypothetical prote
9	35	68.6	2219	T27684	hypothetical prote
10	34	66.7	271	D84695	probable tropinone
11	34	66.7	275	B70640	hypothetical prote
12	34	66.7	748	S59327	hypothetical prote
13	34	66.7	753	T19118	acid phosphatase h
14	34	66.7	1056	S56039	RIC1 protein - yea
15	33	64.7	183	G75378	hypothetical prote
16	33	64.7	271	S62485	probable mitochond
17	33	64.7	278	T00613	probable sugar nuc
18	33	64.7	319	G86832	hypothetical prote
19	33	64.7	426	T31919	hypothetical prote
20	33	64.7	468	F71422	hypothetical prote
21	33	64.7	505	C64502	hypothetical prote
22	33	64.7	584	T09113	response-regulator
23	33	64.7	635	S74718	sulfite reductase
24	33	64.7	707	Q08635	phosphotransferase
25	33	64.7	860	C86203	hypothetical prote
26	33	64.7	1198	T49726	hypothetical prote
27	32	62.7	103	S52493	probable membrane
28	32	62.7	163	C71039	hypothetical prote
29	32	62.7	203	T41226	probable ser-thr p

30 32 62.7 232 2 S32963 hypothetical prote
31 32 62.7 277 2 T23319 hypothetical prote
32 32 62.7 301 2 T45859 R2R3-MYB transcrip
33 32 62.7 304 1 S71285 myb-related protei
34 32 62.7 305 2 T41372 probable ubiquitin
35 32 62.7 346 2 F72672 hypothetical prote
36 32 62.7 424 2 C82516 type I restriction
37 32 62.7 425 1 GLJY14 beta-glucosidase (
38 32 62.7 440 1 TVVPM middle T antigen -
39 32 62.7 489 2 A81166 type II site-speci
40 32 62.7 625 2 T06184 sucrose-fructan 6
41 32 62.7 679 2 T24365 hypothetical prote
42 32 62.7 1019 2 F70342 cation efflux syst
43 32 62.7 1430 2 S50596 hypothetical prote
44 32 62.7 3163 1 JQ1895 genome polyprotein
45 31.5 61.8 652 2 B82724 cardiolipin syntha

ALIGNMENTS

RESULT 1

JE0291
FB19 protein - human
C:Species: Homo sapiens (man)
C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C:Accession: JE0291
R:Totaro, A.; Grifa, A.; Carella, M.; Rommens, J.M.; Valentino, M.A.; Roetto, A.; Zel
Biochem. Biophys. Res. Commun. 250, 555-557, 1998
A:Title: Cloning of a new gene (FB19) within HLA class I region.
A:Reference number: JE0291; MUID:99003493
A:Accession: JE0291
A:Molecule type: mRNA
A:Residues: 1-940 <TOT>
A:Cross-references: GB:V13247; NID:92117158; PIDN:CAA73697.1; PID:92117159
C:Genetics:
A:Gene: FB19
A:Map position: 6p21.3

Query Match 70.6%; Score 36; DB 2; Length 940;
Best Local Similarity 60.0%; Pred. No. 43;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 YSFKPMPLXR 10
I: ||:|||
DB 236 YNLKPIQLKR 245

RESULT 2

T48372
AtB'alpha regulatory subunit of PP2A - Arabidopsis thaliana
N:Alternate names: protein F12E4.240
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48372
R:Bevan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke,
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224492
A:Accession: T48372
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-495 <BEV>
A:Cross-references: EMBL:AL162751
A:Experimental source: cultivar Columbia; DAC clone F12E4

C:Genetics:
A:Map position: 5
A:Introns: 387/3
A:Note: F12E4.240

Query Match 68.6%; Score 35; DB 2; Length 495;
Best Local Similarity 60.0%; Pred. No. 33;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSRKPMPLXR 10
II :I:II I
Db 67 YSVEPLPLFR 76

RESULT 3
T42255
HNF-3/forkhead protein homolog daf-16 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 05-May-2000
C:Accession: T42255
R:Lin, K.; Dorman, J.B.; Rodan, A.; Kenyon, C.
Science 270, 1319-1322, 1997
A:Title: Daf-16, An HNF-3/forkhead family member that can function to double the life span of *C. elegans*
A:Reference number: 222130; MUID:98028757
A:Accession: T42255
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-510 <LIN>
A:Cross-references: EMBL:AF032112; NID:g2623942; PIDN:AAC47803.1; PID:g2623943
C:Genetics:
A:Gene: daf-16

Query Match 68.6%; Score 35; DB 2; Length 510;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
Db 417 NFKPMPL 423

RESULT 4
T42234
fork head-related transcription factor homolog - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jul-2000
C:Accession: T42234
R:Ogg, S.; Paradis, S.; Gottlieb, S.; Patterson, G.I.; Lee, L.; Tissenbaum, H.A.; Ruvkun
Nature 389, 994-999, 1997
A:Title: The Fork head transcription factor DAF-16 transduces insulin-like metabolic and
A:Reference number: 222108; MUID:98013175
A:Accession: T42234
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-530 <OGG>
A:Cross-references: EMBL:AF020344; NID:g2618980; PIDN:AAB84392.1; PID:g2618981
C:Genetics:
A:Gene: daf-16
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

Query Match 68.6%; Score 35; DB 2; Length 530;
Best Local Similarity 85.7%; Pred. No. 36;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
Db 437 NFKPMPL 443

RESULT 5
T37204
hypothetical protein RL3H8.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 02-Jun-2000
C:Accession: T37204
R:Jones, K.; Hinds, K.; Sutterer, C.; Cotman, M.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of *C. elegans* cosmid RL3H8.

A:Reference number: 221633
A:Accession: T37204
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-622 <WIL>
A:Cross-references: EMBL:AF039717; PIDN:AAB96741.1
C:Genetics:
A:Map position: 1
A:Introns: 214/3; 268/3; 329/2; 419/3; 457/2; 547/2; 563/1
A:Note: RL3H8.1
C:Superfamily: unassigned fork head proteins; fork head DNA-binding domain homology

Query Match 68.6%; Score 35; DB 2; Length 622;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
Db 454 NFKPMPL 460

RESULT 6
T23775
hypothetical protein M162.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T23775
R:Basham, V.
submitted to the EMBL Data Library, November 1996
A:Reference number: 219797
A:Accession: T23775
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-690 <WIL>
A:Cross-references: EMBL:X82278; PIDN:CAB05258.1; GSPDB:GN00023; CESP:M162.7
A:Experimental source: clone M162
C:Genetics:
A:Gene: CESP:M162.7
A:Map position: 5
A:Introns: 37/1; 389/2
C:Superfamily: Caenorhabditis elegans hypothetical protein Y116A8C.33

Query Match 68.6%; Score 35; DB 2; Length 690;
Best Local Similarity 85.7%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
Db 323 SFKPLPL 329

RESULT 7
T27066
hypothetical protein Y51A2B.6a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T27066
R:McMurray, A.
submitted to the EMBL Data Library, January 1998
A:Reference number: 220305
A:Accession: T27066
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-717 <WIL>
A:Cross-references: EMBL:AL021493; PIDN:CAA16394.1; GSPDB:GN00023; CESP:Y51A2B.6a
A:Experimental source: clone Y51A2B
C:Genetics:
A:Gene: CESP:Y51A2B.6a
A:Map position: 5
A:Introns: 8/1; 33/1
C:Superfamily: Caenorhabditis elegans hypothetical protein Y116A8C.33

Query Match 68.6%; Score 35; DB 2; Length 717;
 Best Local Similarity 85.7%; Pred. No. 50;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
 |||||I
 Db 309 SFKPLPL 315

RESULT 8
 T27067
 hypothetical protein Y51A2B.6b - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T27067
 R:McMurray, A.
 submitted to the EMBL Data Library, January 1998
 A:Reference number: Z20305
 A:Accession: T27067
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-717 <WII>
 A:Cross-references: EMBL:AL021493; PIDN:CAA16395.1; GSPDB:GN00023; CESP:Y51A2B.6b
 A:Experimental source: clone Y51A2B
 C:Genetics:
 A:Gene: CESP:Y51A2B.6b
 A:Map position: 5
 A:Introns: 8/1; 33/1
 C:Superfamily: Caenorhabditis elegans hypothetical protein Y116A8C.33

Query Match 68.6%; Score 35; DB 2; Length 717;
 Best Local Similarity 85.7%; Pred. No. 50;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
 |||||I
 Db 309 SFKPLPL 315

RESULT 9
 T27684
 hypothetical protein ZK1067.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T27684
 R:Thomas, K.
 submitted to the EMBL Data Library, March 1996
 A:Reference number: Z20404
 A:Accession: T27684
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-2219 <WII>
 A:Cross-references: EMBL:Z70038; PIDN:CAA93884.1; GSPDB:GN00020; CESP:ZK1067.2
 A:Experimental source: clone ZK1067
 C:Genetics:
 A:Gene: CESP:ZK1067.2
 A:Map position: 2
 A:Introns: 139/3; 246/3; 487/3; 784/1; 960/2; 1067/3; 1113/3; 1171/3; 1321/2; 1383/3; 14

Query Match 68.6%; Score 35; DB 2; Length 2219;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 SFKPMPLXR 10
 |||||I
 Db 678 SFKPLPER 686

RESULT 10

084695
 probable tropinone reductase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: D84695
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanNken, S.E.; Umayam, L.; Tallica,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventur
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: D84695
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-271 <STO>
 A:Cross-references: GB:AE002093; MID:g3980399; PIDN:AAC95202.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g29360
 A:Map position: 2
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 66.7%; Score 34; DB 2; Length 271;
 Best Local Similarity 71.4%; Pred. No. 27;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPKPMPL 7
 :|||:I
 Db 265 FSPKPLP 271

RESULT 11
 B70640
 hypothetical protein Rv0687 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: B70640
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: B70640
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-275 <COL>
 A:Cross-references: GB:Z84395; GB:AL123456; MID:g3261698; PIDN:CA806450.1; PID:g18061
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv0687

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 F:11-204/Domain: short-chain alcohol dehydrogenase homology <SADH>

Query Match 66.7%; Score 34; DB 2; Length 275;
 Best Local Similarity 75.0%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKPMPL 8
 :|||I
 Db 228 HSPKPMPL 235

RESULT 12
 S59327
 hypothetical protein YLR135w - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein L3140
 C:Species: Saccharomyces cerevisiae
 C:Date: 29-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 29-Oct-1999
 C:Accession: S59327; S64977
 R:Belius, H.

submitted to the EMBL Data Library, June 1995
A:Description: 36.8 kb of S.cerevisiae chromosome XII including ACE2, CK11, PDC5, SLS1,
A:Reference number: S59313
A:Accession: S59327
A:Molecule type: DNA
A:Residues: 1-748
A:Cross-references: EMBL:X91258; NID:g995686; PIDN:CAA62650.1; PID:g995701
A:Experimental source: strain S288C
R:Delius, H.; Hebling, U.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64967
A:Accession: S64977
A:Molecule type: DNA
A:Residues: 1-748
A:Cross-references: EMBL:Z73307; NID:g1360550; PIDN:CAA97706.1; PID:e245572; PID:g136055
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 12R

Query Match 66.7%; Score 34; DB 2: Length 748;
Best Local Similarity 62.5%; Pred. No. 82;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPKPMPL 8
DB 703 YTFEPIPL 710
I: I: I: I

RESULT 13
T19118
acid phosphatase homolog F26C11.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T19118; T21390
R:Lloyd, C.

submitted to the EMBL Data Library, October 1995
A:Reference number: Z19074
A:Accession: T19118
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-755 <WIL>
A:Cross-references: EMBL:Z54342; PIDN:CAA91156.1; GSPDB:GN00020; CESP:F26C11.1
A:Experimental source: clone C08H9
R:Matthews, P.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19416
A:Accession: T21390
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-755 <W12>
A:Cross-references: EMBL:Z47072; PIDN:CAA87370.1; GSPDB:GN00020; CESP:F26C11.1
A:Experimental source: clone F26C11
C:Genetics:
A:Gene: CESP:F26C11.1
A:Map position: 2
A:Introns: 131/1; 199/3; 277/1; 316/2; 352/3; 417/3; 449/1; 475/3; 577/1; 635/3; 728/3

Query Match 66.7%; Score 34; DB 2: Length 755;
Best Local Similarity 71.4%; Pred. No. 83;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKPMPL 7
DB 718 YQKPIPL 724
I: I: I: I

RESULT 14
S56039
RIC1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L1917; protein YLR039c
C:Species: Saccharomyces cerevisiae

C:Date: 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 20-Jun-2000
C:Accession: S56039; S64866
R:Mizuta, K.
submitted to the EMBL Data Library, December 1994
A:Reference number: S56039
A:Accession: S56039
A:Molecule type: DNA
A:Residues: 1-1056 <M12>
A:Cross-references: EMBL:D43895; NID:g619062; PID:g633123
R:Koetter, P.; Rose, M.; Entlian, K.D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64863
A:Accession: S64866
A:Molecule type: DNA
A:Residues: 1-1056 <KOB>
A:Cross-references: EMBL:Z73211; NID:g1360365; PID:g1360366; MIPS:YLR039c
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:RIC1
A:Cross-references: SGD:S0004029; MIPS:YLR039c
A:Map position: 12R
C:Superfamily: Saccharomyces cerevisiae RIC1 protein

Query Match 66.7%; Score 34; DB 2: Length 1056;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPKPMPL 8
DB 91 YNFKPMAL 98
I: I: I: I

RESULT 15
G75378
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: G75378
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, M.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: G75378
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-183 <WH1>
A:Cross-references: GB:AE002002; GB:AE000513; NID:g6459345; PIDN:AAF11153.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1590
A:Map position: 1

Query Match 64.7%; Score 33; DB 2: Length 183;
Best Local Similarity 71.4%; Pred. No. 27;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPKPMPL 7
DB 99 WNFKPMPL 105
I: I: I: I

Search completed: February 27, 2002, 11:45:15
Job time: 694 sec

Thu Feb 28 09:50:40 2002

us-09-446-109a-4.rpr

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:41 ; Search time 78.39 Seconds
(without alignments)
4.677 Million cell updates/sec

Title: US-09-446-109A-4
Perfect score: 51
Sequence: 1 YSEKPMPLXR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	66.7	755	1 PPAX_CAEEL	Q09549 caenorhabdi
2	34	66.7	1056	1 RIC1_YEAST	P40395 saccharomyc
3	33	64.7	271	1 YAD8_SCHPO	Q09834 schizosacch
4	33	64.7	635	1 CIR_SYN13	P72854 synchocyst
5	33	64.7	707	1 GCVK_HCMVA	P16788 human cytom
6	32	62.7	232	1 YB9H_YEAST	P38340 saccharomyc
7	32	62.7	425	1 BGLT_TRIIP	P26205 trifolium r
8	32	62.7	440	1 TAMI_POVM3	P03076 mouse polyo
9	32	62.7	828	1 PMFC_PROMI	P53514 proteus mir
10	32	62.7	1430	1 YES3_YEAST	P40061 saccharomyc
11	32	62.7	3163	1 POLG_TUMVQ	Q02597 t genome po
12	31	60.8	94	1 SY26_HUMAN	Q9Y258 homo sapien
13	31	60.8	185	1 SODC_FRATU	Q59448 francisella
14	31	60.8	334	1 Y456_MYCGE	P47694 mycoplasma
15	31	60.8	453	1 YKKE_YEAST	P34252 saccharomyc
16	31	60.8	536	1 AYRD_MOUSE	P27040 mus muscucu
17	31	60.8	622	1 SYT_METJA	Q58597 methanococc
18	31	60.8	622	1 COX1_BACSU	P24010 bacillus su
19	31	60.8	863	1 T2C2_RAT	Q9Q281 rattus norv
20	31	60.8	896	1 YFIQ_ECOLI	P76594 escherichia
21	31	60.8	916	1 SCER_LIMPO	Q25386 limulus pol
22	31	60.8	1087	1 PER_DROVI	P12349 drosophila
23	31	60.8	1093	1 TMF1_HUMAN	P82094 homo sapien
24	31	60.8	1250	1 SSD1_YEAST	P24276 saccharomyc
25	31	60.8	1349	1 IK13_YEAST	Q06706 saccharomyc
26	31	60.8	1385	1 YMS5_CAEEL	P34501 caenorhabdi
27	31	60.8	1462	1 TOP2_PEA	Q24308 pisum sativ
28	31	60.8	1701	1 U520_HUMAN	Q75643 homo sapien
29	31	60.8	2148	1 VIT1_AEDAE	Q16927 aedes aegyp
30	31	60.8	3707	1 PGBM_MOUSE	Q05793 mus muscucu
31	31	60.8	4393	1 PGBM_HUMAN	P98160 homo sapien
32	30	58.8	93	1 YARO_ERWIE	P42518 erwinia her
33	30	58.8	177	1 SODC_SALTY	P53636 salmonella

34	30	58.8	181	1 Y4AS_RHISN	P55366 rhizobium s
35	30	58.8	239	1 Y247_MYCPN	P75428 mycoplasma
36	30	58.8	289	1 ATP6_PROMO	P21903 propionigen
37	30	58.8	356	1 TRM6_AQUAE	O67274 aquifex aeo
38	30	58.8	402	1 LAG2_CAEEL	P45442 caenorhabdi
39	30	58.8	417	1 SYS_MYCGE	P47251 mycoplasma
40	30	58.8	439	1 GFO_ZYMMO	Q07982 zymomonas m
41	30	58.8	440	1 SCAL_DROME	P30052 drosophila
42	30	58.8	477	1 YMN2_CAEEL	P54246 caenorhabdi
43	30	58.8	486	1 RBL1_RHOSH	P27997 rhodobacter
44	30	58.8	487	1 EBN2_EBV	P12978 epstein-bar
45	30	58.8	554	1 MHPA_ECOLI	P77397 escherichia

ALIGNMENTS

RESULT 1	PPAX_CAEEL	STANDARD	PRT: 755 AA.
ID	Q09549; Q17843;		
AC	01-NOV-1995 (Rel. 32, Created)		
DT	30-MAY-2000 (Rel. 39, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	PUTATIVE ACID PHOSPHATASE F26C11.1 (EC 3.1.3.2).		
GN	F26C11.1.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRATN=BRISTOL N2;		
RA	Matthews P., Lloyd C.;		
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.		
CC	-!- CATALYTIC ACTIVITY: AN ORTHOPHOSPHORIC MONOESTER + H(2)O -> AN		
CC	ALCOHOL + ORTHOPHOSPHATE.		
CC	-!- SIMILARITY: BELONGS TO THE HISTIDINE ACID PHOSPHATASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement. (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; 247072; CAA87370.1;		
DR	EMBL; 254342; CAA87370.1; JOINED.		
DR	EMBL; 254342; CAA91156.1; -.		
DR	EMBL; 247072; CAA91156.1; JOINED.		
DR	WormPep; F26C11.1; CF05732.		
DR	InterPro; IPR000560; His_acid_phosphatase.		
DR	Pfam; PF00328; acid_phosphat_1.		
DR	PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.		
DR	PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.		
DR	Hypothetical protein; Hydrolase.		
FT	ACT_SITE 356 356 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).		
FT	ACT_SITE 399 399 BY SIMILARITY.		
FT	ACT_SITE 636 636 PROTON DONOR (BY SIMILARITY).		
FT	DISULFID 702 708 BY SIMILARITY.		
SQ	SEQUENCE 755 AA; 87783 MW; E7F3050115A235EB CRC64;		

Query Match 66.7%; Score 34; DB 1; Length 755;
Best Local Similarity 71.4%; Pred. No. 38;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSEKPMPL 7
| | | | |
Db 718 YQFKPLP 724

```
RESULT 2
RIC1_YEAST
ID RIC1_YEAST STANDARD; PRT: 1056 AA.
AC P40395;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE RIC1 PROTEIN.
GN RIC1 OR YLR039C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Mizuta K.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Koetter P., Rose M., Entian K.D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE TRANSCRIPTION OF BOTH RIBOSOMAL PROTEIN
CC GENES AND RIBOSOMAL RNA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D43895; BAA07866.1; -
CC EMBL: 273211; CAA97567.1; -
CC SGD: S0004029; RIC1.
SQ SEQUENCE 1056 AA; 121649 MW; 2368616867AEAU31 CRC64;

Query Match 66.7%; Score 34; DB 1; Length 1056;
Best Local Similarity 75.0%; Pred. No. 53;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSKPMPPL 8
DB 91 YNFKPMAL 98

RESULT 3
YADB_SCHPO
ID YADB_SCHPO STANDARD; PRT: 271 AA.
AC Q09834;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE PUTATIVE MITOCHONDRIAL CARRIER C4G8.08.
GN SPAC48.08.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Strain=972;
RA Badcock K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (POTENTIAL).
CC -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90901; BAA16869.1; -
CC HSSP: P17846; 7GEP.
CC InterPro: IPR000660; Nlr_Sir.
CC Pfam: PF01077; Nlr_Sir; 1.

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 256276; CAA91209.1; -
CC InterPro: IPR001993; Mitoch_Carrier.
CC Pfam: PF00153; mitoch_Carrier; 3.
CC PROSITE: PS00215; MITOCH_CARRIER; 1.
CC KW Hypothetical protein; Mitochondrion; Inner membrane; Repeat;
CC Transmembrane; Transport.
CC FT TRANSMEM 49 69 POTENTIAL.
CC FT TRANSMEM 84 104 POTENTIAL.
CC FT TRANSMEM 138 158 POTENTIAL.
CC FT TRANSMEM 170 190 POTENTIAL.
SQ SEQUENCE 271 AA; 30002 MW; EF330C8C8BDB6C4 CRC64;

Query Match 64.7%; Score 33; DB 1; Length 271;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSKPMPPL 8
DB 36 FSKTWPPL 43

RESULT 4
SIR_SYNY3
ID SIR_SYNY3 STANDARD; PRT: 635 AA.
AC P72854;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SULFITE REDUCTASE (FERREDOXIN) (EC 1.8.7.1).
GN SIR OR SLR0963.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- CATALYTIC ACTIVITY: H(2)S + 3 OXIDIZED FERREDOXIN + 3 H(2)O =
CC SULFITE + 3 REDUCED FERREDOXIN
CC -!- COFACTOR: THIS ENZYME CONTAINS ONE STROHEME AND ONE 4FE-4S
CC IRON-SULFUR CENTER AS PROSTHETIC GROUPS (HY SIMILARITY).
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/STROHEME DOMAIN
CC FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
CC SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D90901; BAA16869.1; -
CC HSSP: P17846; 7GEP.
CC InterPro: IPR000660; Nlr_Sir.
CC Pfam: PF01077; Nlr_Sir; 1.
```

DR PRINTS: PK00397; STROHAEM.
 DR PROSITE; PS00365; NTR_SIR; 1.
 KW Oxidoreductase; Heme; Iron-sulfur; 4Fe-4S; Complete proteome.
 FT METAL 451 451 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 457 457 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 496 496 IRON-SULFUR (4FE-4S) (POTENTIAL).
 FT METAL 500 500 IRON-SULFUR (4FE-4S) AND SIROHEME
 (BY SIMILARITY).
 SQ SEQUENCE 635 AA; 71441 MW; 91F4A1B139AD7B34 CRC64;

Query Match: 64.7%; Score 33; DB 1; Length 635;
 Rest local Similarity 66.7%; Pred. No. 49;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 SFKPMPLXR 10
 I I I I I I I
 DB 347 SFKPLPAWR 355

RESULT 5
 GCVK_HCMVA STANDARD; PRT: 707 AA.
 AC P16788;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE GANCICLOVIR KINASE (EC 2.7.1.-) (HSR3 PROTEIN).
 GN UL97.
 OS Human cytomegalovirus (strain AD169).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Cytomegalovirus.
 OX NCBI_TaxID=10360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9026039; PubMed=2161319;
 RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
 RA Horsnell T., Hutchinson C.A. III, Kozarides T., Martignetti J.A.,
 RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.;
 RT "Analysis of the protein-coding content of the sequence of human
 cytomegalovirus strain AD169."
 RL Curr. Top. Microbiol. Immunol. 154:125-169(1990).
 RN [2]
 RP FUNCTION.
 RX MEDLINE=92310592; PubMed=1319559;
 RA Littler E., Stuart A.D., Chee M.S.;
 RT "Human cytomegalovirus UL97 open reading frame encodes a protein that
 phosphorylates the antiviral nucleoside analogue ganciclovir";
 RL Nature 358:160-162(1992).
 RN [3]
 RP FUNCTION.
 RX MEDLINE=92310593; PubMed=1319560;
 RA Sullivan V., Talarico C.L., Stanat S.C., Davis M., Coen D.M.,
 RA Biron K.K.;
 RT "A protein kinase homologue controls phosphorylation of ganciclovir
 in human cytomegalovirus-infected cells."
 RL Nature 358:162-164(1992).
 RN [4]
 RP ERRATUM.
 RX MEDLINE=92396223; PubMed=1326083;
 RA Sullivan V., Talarico C.L., Stanat S.C., Davis M., Coen D.M.,
 RA Biron K.K.;
 RL Nature 359:85-85(1992).
 RN [5]
 RP ERRATUM.
 RA Sullivan V., Talarico C.L., Stanat S.C., Davis M., Coen D.M.,
 RA Biron K.K.;
 HL Nature 366:756-756(1993).
 CC -!- FUNCTION: PHOSPHORYLATES THE ANTIVIRAL NUCLEOSIDE ANALOG
 GANCICLOVIR.
 CC -!- SIMILARITY: BELONGS TO THE HCMV UL97 / HVS 36 FAMILY.
 CC -!- SIMILARITY: TO TYROSINE-PROTEIN KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X17403; CAA35333.1; -
 DR PIR: S09862; QOBEJ5.
 DR HSP: P06722; IAZO.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR001245; Tyr_kin.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS00111; PROTEIN_KINASE_DOM; FALSE_NEG.
 KW Transferase; Kinase; ATP-binding.
 FT DOMAIN 41 63 ALA-RICH
 FT NP_BIND 337 345 ATP (BY SIMILARITY).
 FT BINDING 359 359 ATP (BY SIMILARITY).
 FT ACT_SITE 456 456 BY SIMILARITY.
 SQ SEQUENCE 707 AA; 78232 MW; 74914183E5A5E03A CRC64;

Query Match: 64.7%; Score 33; DB 1; Length 707;
 Rest local Similarity 55.6%; Pred. No. 55;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 SFKPMPLXR 10
 I I I I I I I
 DB 522 AFRPMLQK 530

RESULT 6
 YB9H_YEAST STANDARD; PRT: 232 AA.
 AC P38340;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE HYPOTHETICAL 26.1 KDA PROTEIN IN POP4-SHM1 INTERGENIC REGION.
 GN YBR261C OR YBK1729.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=93220397; PubMed=8465606;
 RA Dolignon F., Bileau N., Crouzet M., Algle M.;
 RT "The complete sequence of a 19,482 bp segment located on the right
 arm of chromosome II from Saccharomyces cerevisiae."
 RL Yeast 9:189-199(1993).
 CC -!- SIMILARITY: TO S.POMME SPAC1688.14C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X70529; CAA49926.1; -
 DR EMBL: X36130; CAA85224.1; -
 DR PIR: S32963; S32963.
 DR SGD: S0000465; YBR261C.
 KW Hypothetical protein.
 SQ SEQUENCE 232 AA; 26068 MW; 66699F37B0013088 CRC64;

Query Match: 62.7%; Score 32; DB 1; Length 232;

Best Local Similarity 71.4%; Pred. No. 27;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSRKPMPL 7
Db 225 YALKPMP 231

RESULT 7

BGLT_TRIP
ID BGLT_TRIP STANDARD; PRT; 425 AA.
AC P26205;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CYANOGENIC BETA-GLUCOSIDASE PRECURSOR (EC 3.2.1.21) (LINMARASE)
DE (FRAGMENT).
GN Li.

OS Trifolium repens (Creeping white clover).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.
OX NCBI_TaxID=3899;
RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 12-25 AND 125-147.

RC STRAIN=S100 (EG); TISSUE=leaf;

RX MEDLINE=91322517; PubMed=1907511;

RA Oxtoby E., Dunn M.A., Pancoro A., Hughes M.A.;

RT "Nucleotide and derived amino acid sequence of the cyanogenic beta-glucosidase (linamarase) from white clover (Trifolium repens L.);
Plant Mol. Biol. 17:209-219(1991).
RL [2]

RN X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS).

RP MEDLINE=96097401; PubMed=8535788;

RA Barrett T., Suresh C.G., Tolley S.P., Dodson E.J., Hughes M.A.;

RT "The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase.";

RL Structure 3:951-960(1995)

CC -1- FUNCTION: HYDROLYSES CYANOGLUCOSIDES, CONTRIBUTING TO THE RELEASE

OF HYDROXYANIC ACID, WHICH FUNCTIONS AS A DEFENSE MECHANISM

AGAINST SMALL PREDATORS, WHEN THE LEAF TISSUE IS DAMAGED.

CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-

GLUCOSE RESIDUES WITH RELEASE OF BETA-D-GLUCOSE.

CC -1- PATHWAY: CYANOGENESIS.

CC -1- SUBUNIT: HOMODIMER.

CC -1- TISSUE SPECIFICITY: LEAVES.

CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF GLYCOSYL HYDROLASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X56733; CAA40057.1; -

DR PIR: S16580; GLJY14.

DR PDB: 1CBG; 15-OCT-95.

DR InterPro: IPR001360; Glyco_hydro_1.

DR Pfam: PF00232; Glyco_hydro_1; 1.

DR PROSITE: PS00572; GLYCOSYL-HYDROL_FL1; 1.

DR PROSITE: PS00653; GLYCOSYL-HYDROL_FL2; 1.

KW Hydrolase: Glycosidase; Glycoprotein; Signal; 3D-structure.

FT NON_TER 1 1

FT SIGNAL <1 11

FT CHAIN 12 425 CYANOGENIC BETA-GLUCOSIDASE.

FT ACT_SITE 194 194 PROTON DONOR.

FT ACT_SITE 408 408 NUCLEOPHILE.

FT DISULFID 213 221

FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 425 AA; 48324 MW; 20B983B65C47A678 CRC64;

Query Match 62.7%; Score 32; DB 1; Length 425;
Best Local Similarity 50.0%; Pred. No. 51;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSRKPMPL 8
Db 10 HAKRPLPI 17

RESULT 8

TAMI_POWM3
ID TAMI_POWM3 STANDARD; PRT; 440 AA.

AC P03076;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MIDDLE T ANTIGEN.

OS Mouse polyomavirus (strain 3).

OC Viruses; dsDNA viruses, no RNA stage; Polyomaviridae; Polyomavirus.

OX NCBI_TaxID=10635;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=81053800; PubMed=6253927;

RA Deininger P.L., Esty A., Laporte P., Hsu H., Friedmann T.;

RT "The nucleotide sequence and restriction enzyme sites of the polyoma genome.";

RL Nucleic Acids Res. 8:855-860(1980).

CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.

DR PIR: A03613; TVVPM

DR InterPro: IPR001623; DnaJ_N.

DR InterPro: IPR003354; Papo_T_antigen.

DR Pfam: PF00226; DnaJ_1.

DR Pfam: PF02380; Papo_T_antigen; 1.

DR SMART: SM00271; DnaJ_1.

DR PROSITE: PS00636; DnaJ_1; FALSE_NEG.

DR PROSITE: PS00076; DnaJ_2; FALSE_NEG.

KW Early protein.

FT DOMAIN 12 75 J-DOMAIN.

SQ SEQUENCE 440 AA; 50831 MW; 63A8E82E532BCC37 CRC64;

Query Match 62.7%; Score 32; DB 1; Length 440;
Best Local Similarity 50.0%; Pred. No. 53;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSRKPMPLXR 10
Db 203 YSYRALPIER 212

RESULT 9

PMFC_PROMI
ID PMFC_PROMI STANDARD; PRT; 828 AA.

AC P33514;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE OUTER MEMBRANE USHER PROTEIN PMFC PRECURSOR.

GN PMFC

OS Proteus mirabilis.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Proteus.

OX NCBI_TaxID=584;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=H4320;

RX MEDLINE=95047519; PubMed=7959033;

RA Massad G., Mobley H.L.T.;

RT "Genetic organization and complete sequence of the proteus mirabilis

pmf fimbrial operon.";

RL Gene 150:101-104(1994).

CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF PMF FIMBRIAL
 CC SUBUNITS ACROSS THE OUTER MEMBRANE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
 CC (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z35428; CAA84590.1; -;
 DR InterPro: IPR000015; Fimb_usher.
 DR Pfam: PF00577; Usher; 1.
 DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
 KW Outer membrane; Transmembrane; Fimbria; Transport; Signal.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 828 OUTER MEMBRANE USHER PROTEIN PMFC.
 SQ SEQUENCE 828 AA: 93108 MW: 111805992A64BA44 CRC64;
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

Query Match 62.7%; Score 32; DB 1; Length 828;
 Best Local Similarity 57.1%; Pred. No. 1e+02;
 Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YSKPKMP 7
 DB 228 YAYKPLP 234
 I:|:|:|
 I:|:|:|

RESULT 10
 YES3_YEAST
 ID YES3_YEAST STANDARD; PRT: 1430 AA.
 AC P40061;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HYPOTHETICAL 164.4 KDA PROTEIN IN MET6-PUP3 INTERGENIC REGION.
 GN YER093C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Guzman E., Hartzell G., Hunkeler-Smith S.,
 RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
 RA Patel F.X., Roberts D., Schl P., Schramm S., Shogren T., Smith V.,
 RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 967-1430 FROM N.A.
 RC STRAIN=S288C;
 RA Korch C., Mountain H.A., Wenzlau J.M.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO S POMBE STE16
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U18839; AAB64648.1; -;
 DR EMBL: U32508; AAB60298.1; -;
 DR SGD: S0000895; YER093C.
 DR InterPro: IPR000651; RasGEFN.
 DR Pfam: PF00618; RasGEFN; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 380 400 POTENTIAL.
 FT TRANSMEM 565 585 POTENTIAL.
 FT TRANSMEM 620 640 POTENTIAL.
 FT TRANSMEM 870 890 POTENTIAL.
 FT TRANSMEM 1149 1169 POTENTIAL.
 FT TRANSMEM 1181 1201 POTENTIAL.
 SQ SEQUENCE 1430 AA: 164367 MW: 07BB5D75BA204E50 CRC64;
 CC -----
 CC Query Match 62.7%; Score 32; DB 1; Length 1430;
 CC Best Local Similarity 71.4%; Pred. No. 1.8e+02;
 CC Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 YSKPKMP 7
 DB 600 HSEKPIP 606
 I:|:|:|
 I:|:|:|

RESULT 11
 POLG_TUMVQ
 ID POLG_TUMVQ STANDARD; PRT: 3163 AA.
 AC G02597;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GENOME POLYPROTEIN [CONTAINS: N-TERMINAL PROTEIN (PI); HELPER
 DE COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN
 DE 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (6K2);
 DE GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA)
 DE (EC 3.4.22.44) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION
 DE PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48);
 DE COAT PROTEIN (CP)].
 OS Turnip mosaic virus (strain Quebec) (TumV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
 OC Potyvirus.
 OX NCBI_TaxID=36396;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93057350; PubMed=1431807;
 RA Nicolas O., Laliberte J.F.;
 RT "The complete nucleotide sequence of turnip mosaic potyvirus RNA.";
 RL J. Gen. Virol. 73:2785-2793(1992).
 RN [2]
 RP SEQUENCE OF 1534-3163 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=91073142; PubMed=2254757;
 RA Tremblay M.F., Nicolas O., Sinha R., Lazure C., Laliberte J.F.;
 RT "Sequence of the 3'-terminal region of turnip mosaic virus RNA and
 RT the capsid protein gene.";
 RL J. Gen. Virol. 71:2769-2772(1990).
 CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
 CC MAY BE INVOLVED IN REPLICATION.
 CC -1- CATALYTIC ACTIVITY: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: NUCLEAR INCLUSION PROTEIN A IS FURTHER RESTRICTED BY
 CC GLUTAMINYL BONDS, AND ACTIVITY IS FURTHER RESTRICTED BY
 CC PREFERENCES FOR THE AMINO ACIDS IN P6 -PI THAT VARY WITH THE
 CC SPECIES OF POTYVIRUS, E.G. GLU-XAA-TYK-XAA-GIN+(SER OR GLY)
 CC FOR THE ENZYME FROM TOBACCO ETCH VIRUS. THE NATURAL SUBSTRATE IS
 CC THE VIRAL POLYPROTEIN, BUT OTHER PROTEINS AND OLIGOPEPTIDES
 CC CONTAINING THE APPROPRIATE CONSENSUS SEQUENCE ARE ALSO CLEAVED.
 CC -1- PFM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
 CC -1- PFM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
 CC POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
 CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
 CC INDIVIDUAL PROTEINS.
 CC -1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
 CC -1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
 CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL: D10927; BAA01725.1; -;
 DR EMBL: D10601; BAA01452.1; -;
 DR PIR: J01168; J01168.
 DR PIR: J03895; J01895.
 DR MEROPS: C04.001; -;
 DR MEROPS: C06.001; -;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR001730; Peptidase_C4.
 DR InterPro: IPR001456; Peptidase_C6.
 DR InterPro: IPR002540; Poty_P1.
 DR InterPro: IPR001592; Poty_coat.
 DR InterPro: IPR001205; RNA_pol_P3D.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00863; Peptidase_C4; 1.
 DR Pfam: PF00851; Peptidase_C6; 1.
 DR Pfam: PF00767; Poty_coat; 1.
 DR Pfam: PF01577; Poty_P1; 1.
 DR Pfam: PF00680; RNA_dep_RNA_pol; 1.
 DR PRINTS: PR00966; NIAPOTYPTASE.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELICC; 1.
 DR Hydrolase: Transferase; Thiol protease; RNA-directed RNA polymerase;
 KW Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
 KW ATP-binding.
 FT CHAIN 1 362 N-TERMINAL PROTEIN.
 FT CHAIN 363 820 HELPER COMPONENT PROTEINASE.
 FT CHAIN 821 1175 PROTEIN P3.
 FT CHAIN 1176 1227 6 KDA PROTEIN 1.
 FT CHAIN 1228 1870 CYTOPLASMIC INCLUSION PROTEIN.
 FT CHAIN 1871 1923 6 KDA PROTEIN 2.
 FT CHAIN 1924 2115 GENOME-LINKED PROTEIN.
 FT CHAIN 2116 2358 NUCLEAR INCLUSION PROTEIN A.
 FT CHAIN 2359 2875 NUCLEAR INCLUSION PROTEIN B.
 FT CHAIN 2876 3163 COAT PROTEIN.
 FT BINDING 1986 COVALENT LINKAGE OF VIRAL RNA (BY
 FT NP_BIND 1313 1320 ATP (POTENTIAL).
 FT CONFLICT 2862 2862 E -> G (IN REF. 2).
 SQ SEQUENCE 3163 AA; 357817 MW; 6180F73B58DF6D59 CRC64;

Query Match 62.7%; Score 32; DB 1; Length 3163;
 Best Local Similarity 85.7%; Pred. No. 4.2e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Caps 0;

OY 2 SFKPMPL 8
 DB 28 SFPPMPL 34

RESULT 12
 SY26_HUMAN STANDARD; PRT; 94 AA.
 AC Q9Y258;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SMALL INDUCIBLE CYTOKINE A26 PRECURSOR (EOTAXIN-3) (MACROPHAGE
 DE INFLAMMATORY PROTEIN 4-ALPHA) (MIP-4-ALPHA) (CCL26).
 GN SCYA26.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=99303563; PubMed=10373330;
 RA Guo R.F., Ward P.A., Hu S.M., McDuffie J.E., Huber-Lang M., Shi M.M.;
 RT "Molecular cloning and characterization of a novel human CC chemokine,
 RL SCYA26.";
 RL Genomics 58:313-317(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=99343775; PubMed=10415065;
 RA Shinkai A., Yoshisue H., Koike M., Shoji E., Nakagawa S., Saito A.,
 RA Takeda T., Imabeppu S., Kato Y., Hanai N., Anazawa H., Kuga T.,
 RA Nishi T.;
 RT "A novel human CC chemokine, eotaxin-3, that is expressed in IL-4-
 RT stimulated vascular endothelial cells, exhibits potent activity toward
 RT eosinophils.";
 RL J. Immunol. 163:1602-1610(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Jones K., Graves T., Duckels G., Fronick W.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.
 RC TISSUE=Lung;
 RX MEDLINE=99419089; PubMed=10488147;
 RA Kitaura M., Suzuki N., Imai T., Takagi S., Suzuki R., Nakajima T.,
 RA Hirai K., Nomiya H., Yoshie O.;
 RT "Molecular cloning of a novel human CC chemokine (Eotaxin-3) that is a
 RT functional ligand of CC chemokine receptor 3.";
 RL J. Biol. Chem. 274:27975-27980(1999).
 CC -!- FUNCTION: CHEMOTACTIVE FOR EOSINOPHILS AND BASOPHILS. BINDS TO
 CC CCR3.
 CC -!- SUBCELLULAR LOCATION: SECRETED.
 CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED AT LOW LEVELS IN
 CC VARIOUS TISSUES INCLUDING HEART AND OVARY.
 CC -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE
 CC C-C) (CHEMOKINE CC).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>
 CC or send an email to license@lsb-sib.ch).
 CC -----
 DR EMBL: AF124601; AAD22197.1; -;
 DR EMBL: AH016542; BAA36704.1; -;
 DR EMBL: AC005102; AAD15411.1; -;
 DR EMBL: AB010447; BAA84579.1; -;
 DR MIM: 604697; -;
 DR InterPro: IPR001811; Chemokine_IL8.
 DR InterPro: IPR000827; Small_cytokine_CC.
 DR Pfam: PF00048; IL8; 1.
 DR SMART: SM00199; SCY; 1.
 DR PROSITE: PS00472; SMALL_CYTOKINES_CC; 1.
 KW Cytokine; Chemotaxis; Inflammatory response; Signal.
 FT SIGNAL 1 23
 FT CHAIN 24 94 SMALL INDUCIBLE CYTOKINE A26.
 FT DISULFID 33 57 BY SIMILARITY.
 FT DISULFID 34 73 BY SIMILARITY.
 SQ SEQUENCE 94 AA; 10648 MW; 8525B1E4BDD39A5F CRC64;

Query Match 60.8%; Score 31; DB 1; Length 94;
 Best Local Similarity 71.4%; Pred. No. 16;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 YSEKPM 7

```
Db 37 YSHKPLP 43

RESULT 13
SODC_FRATU STANDARD; PRT; 185 AA.
AC Q59448;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE SUPEROXIDE DISMUTASE [CU-ZN] PRECURSOR (EC 1.15.1.1).
GN SODC.
OS Francisella tularensis.
OC Bacteria; Proteobacteria; gamma subdivision; Francisella group;
OC Francisella.
OX NCBI_TaxID=263;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LVS;
RA Clairoux N., Nano F.E., Boissinot M.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN THE
CC CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) -> O(2) + H(2)O(2).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U35670; AAA99764.1; -
CC InterPro: IPR001424; SOD_CU_ZN.
CC Pfam: PF00080; sodcu; 1.
CC DR PROSITE: PD000469; SOD_CU_ZN.1.
CC DR PROSITE: PS00087; SOD_CU_ZN.1; 1.
CC DR PROSITE: PS00332; SOD_CU_ZN.2; FALSE_NEG.
KW Oxidoreductase; Copper; Zinc; periplasmic; Signal.
FT SIGNAL 1 18
FT CHAIN 19 185 SUPEROXIDE DISMUTASE [CU-ZN].
FT METAL 85 85 COPPER (BY SIMILARITY).
FT METAL 87 87 COPPER (BY SIMILARITY).
FT METAL 102 102 COPPER AND ZINC (BY SIMILARITY).
FT METAL 111 111 ZINC (BY SIMILARITY).
FT METAL 120 120 ZINC (BY SIMILARITY).
FT METAL 123 123 ZINC (BY SIMILARITY).
FT METAL 158 158 COPPER (BY SIMILARITY).
FT DISULFID 92 180 BY SIMILARITY.
FT DOMAIN 172 175 POLY-GLY.
SQ SEQUENCE 185 AA: 19887 MW: 730892A80F54403B CRC64;

Query Match 60.8%; Score 31; DB 1; Length 185;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 YSFKPMP 8
DB 164 YSDKQPLP 171

RESULT 14
Y456_MYCCE
ID Y456_MYCCE STANDARD; PRT; 334 AA.
AC P47694; O49238;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

Dt 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN MG456.
GN MG456.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek B.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Brott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 86-182 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek B.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Brott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 86-182 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek B.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Brott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing."
RL J. Bacteriol. 175:7918-7930(1993).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U39727; AAC72476.1; -
CC EMBL: U39732; AAB01646.1; -
CC DR EMBL: U01790; AAD10612.1; -
CC DR TIGR: MG456; -
CC KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSNEM 19 39 POTENTIAL.
FT TRANSNEM 55 75 POTENTIAL.
FT CONFLICT 181 181 F -> S (IN REF..2).
SQ SEQUENCE 334 AA: 38354 MW: D142C358CE474C71 CRC64;

Query Match 60.8%; Score 31; DB 1; Length 334;
Best Local Similarity 57.1%; Pred. No. 63;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSFKPMP 7
DB 128 YNFEPLP 134

RESULT 15
YKK8_YEAST
ID YKK8_YEAST STANDARD; PRT; 453 AA.
AC P34252;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE HYPOTHETICAL 52.3 KDA PROTEIN IN HAP4-AAT1 INTERGENIC REGION.
GN YKL108W OR YKL463.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94152173; PubMed-8109175;
```

RA Cheret G., Pallier C., Valens M., Daignan-Fornier B., Fukuhara H.,
RA Bolotin-Fukuhara M., Sor E.;
RT "The DNA sequence analysis of the HAP4-LAP4 region on chromosome XI
RT of Saccharomyces cerevisiae suggests the presence of a second
RT aspartate aminotransferase gene in yeast.";
RL Yeast 9:1259-1265(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X71133; CAA50449.1; -;
DR EMBL; Z28108; CAA81948.1; -;
DR PIR; S37935; S37935;
DR PIR; S39096; S39096;
DR SGD; S0001591; YKL108W.
KW Hypothetical protein.
SO SEQUENCE 453 AA; 52271 MW; 5282969C97AAEAC8 CRC64;

Query Match 60.8%; Score 31; DB 1; Length 453;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 SFKPMPLXR 10
||| |||
Db 252 SFSPSPILIR 260

Search completed: February 27, 2002, 11:42:42
Job time: 541 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:03 ; Search time 281.76 Seconds
(without alignments)
5.191 Million cell updates/sec

Title: US-09-446-109a-4
Perfect score: 51
Sequence: 1 YSFKPMLXR 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mic:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	72.5	370	Q9W473	Q9W473 drosophila
2	37	72.5	510	Q9A39	Q9A39 trichodesmi
3	36	70.6	872	055000	055000 rattus norv
4	36	70.6	940	000405	000405 homo sapien
5	35	68.6	67	Q38812	Q38812 arabidopsis
6	35	68.6	495	004375	004375 arabidopsis
7	35	68.6	504	065218	065218 arabidopsis
8	35	68.6	504	09S702	09S702 arabidopsis
9	35	68.6	508	016849	016849 caenorhabdi
10	35	68.6	510	016876	016876 caenorhabdi
11	35	68.6	530	016850	016850 caenorhabdi
12	35	68.6	622	044756	044756 caenorhabdi
13	35	68.6	690	Q9XUC6	Q9XUC6 caenorhabdi
14	35	68.6	717	045958	045958 caenorhabdi
15	35	68.6	717	045959	045959 caenorhabdi
16	35	68.6	774	Q9H8B9	Q9H8B9 homo sapien
17	35	68.6	829	P87584	P87584 chipmunk pa
18	35	68.6	939	Q9H088	Q9H088 homo sapien
19	35	68.6	2219	Q23388	Q23388 caenorhabdi

20	34	66.7	271	10	Q9ZW19	Q9ZW19 arabidopsis
21	34	66.7	275	2	P95033	P95033 mycobacteri
22	34	66.7	748	3	Q12098	Q12098 saccharomyc
23	34	66.7	1199	12	Q83066	Q83066 fish lympho
24	34	66.7	2207	5	Q9U0V2	Q9U0V2 leishmania
25	33	64.7	183	2	Q9RU06	Q9RU06 deinococcus
26	33	64.7	202	12	Q9YKH0	Q9YKH0 human cytom
27	33	64.7	205	12	Q9WIL6	Q9WIL6 human cytom
28	33	64.7	206	12	Q9YKG9	Q9YKG9 human cytom
29	33	64.7	206	12	Q9W8X5	Q9W8X5 human cytom
30	33	64.7	218	5	Q9N4M6	Q9N4M6 caenorhabdi
31	33	64.7	278	10	Q64726	Q64726 arabidopsis
32	33	64.7	319	2	Q9CF18	Q9CF18 lactococcus
33	33	64.7	426	5	Q16536	Q16536 caenorhabdi
34	33	64.7	468	10	Q23425	Q23425 arabidopsis
35	33	64.7	505	1	Q59015	Q59015 methanococc
36	33	64.7	584	2	Q68118	Q68118 vibrio chol
37	33	64.7	610	5	Q9N505	Q9N505 caenorhabdi
38	33	64.7	612	12	Q9DWA3	Q9DWA3 rat cytoeq
39	33	64.7	703	10	Q9AVP4	Q9AVP4 nicotiana t
40	33	64.7	706	10	Q9LW81	Q9LW81 arabidopsis
41	33	64.7	707	12	Q68101	Q68101 human cytom
42	33	64.7	707	12	Q68102	Q68102 human cytom
43	33	64.7	739	10	Q9ARD0	Q9ARD0 lycopersico
44	33	64.7	739	10	Q9ARC7	Q9ARC7 lycopersico
45	33	64.7	854	5	Q09238	Q09238 pseudocorti

ALIGNMENTS

RESULT 1

Q9W473 ID Q9W473 PRELIMINARY: PRT: 370 AA.
AC Q9W473;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
UT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG15767 PROTEIN.
GN CG15767
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Eudopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.N., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Duquennois S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira J.A., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houlston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei R., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003435; AAF46085.1;
DR FlyBase: FBgn0029809; CG15767.
DR InterPro: IPR002130; CSA_PPIase.
DR Pfam: PF00160; pro_isomerase; 1.
DR PROSITE: PS00072; CSA_PPIASE_2; 1.
SQ SEQUENCE 370 AA; 42460 MW; 388038EC1BA14F71 CRC64;

Query Match 72.5%; Score 37; DB 5; Length 370;
Best Local Similarity 66.7%; Pred. No. 28;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 SFKPMPLXR 10

Db 321 SFKPLPVAR 329

RESULT 2

ID Q9RA39 PRELIMINARY; PRT: 510 AA.
AC Q9RA39;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FERREDOXIN NITRITE REDUCTASE.
CN NIRA.
OS Trichodesmium sp. WH9601.
OC Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX NCBI_TaxID=104582;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WH9601;
RA Li H., Post A.F.;
RT "The nitrate assimilation genes of the diazotrophic, bloom-forming
cyanobacterium *Trichodesmium* strain WH9601.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF178846; AAF00916.1;
DR InterPro: IPR000660; Nir_Sir.
DR Pfam: PF01077; NIR_SIR; 1.
DR PRINTS: PR00397; SIROHAEM.
DR PROSITE: PS00365; NIR_SIR; 1.
SQ SEQUENCE 510 AA; 57584 MW; F3ED1B8E0BC8619F CRC64;

Query Match 72.5%; Score 37; DB 2; Length 510;
Best Local Similarity 60.0%; Pred. No. 39;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSFKPMPLXR 10

Db 382 FSFEPQLTR 391

RESULT 3

ID Q55000 PRELIMINARY; PRT: 872 AA.
AC Q55000;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHELICAL 92.8 KDA PROTEIN.
CN PNUTS.
OS *Rattus norvegicus* (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RA Allen P.B., Oulmet C.C., Greengard P.;
RL J. Biol. Chem. 0:0-0(1998).
DR EMBL: AF040954; AAB96775.1;
DR InterPro: IPR000571; ZF-CCCH.
DR InterPro: IPR003617; TFS2_N.
DR SMART: SM00509; TFS2N; 1.
DR SMART: SM00356; ZnF_C3H1; 1.
KW Hypothetical protein.
SQ SEQUENCE 872 AA; 92828 MW; 23CC61B4B296E948 CRC64;

Query Match 70.6%; Score 36; DB 11; Length 872;
Best Local Similarity 60.0%; Pred. No. 1e+02;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSFKPMPLXR 10

Db 236 YNLKPIPLKR 245

RESULT 4

ID Q00405 PRELIMINARY; PRT: 940 AA.
AC Q00405;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FB19 PROTEIN.
CN FB19.
OS *Homo sapiens* (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Gasparini P.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y13247; CAA73697.1;
DR InterPro: IPR000571; ZF-CCCH.
DR InterPro: IPR003617; TFS2_N.
DR SMART: SM00509; TFS2N; 1.
DR SMART: SM00356; ZnF_C3H1; 1.
SQ SEQUENCE 940 AA; 99105 MW; D2350ACD5342B024 CRC64;

Query Match 70.6%; Score 36; DB 4; Length 940;
Best Local Similarity 60.0%; Pred. No. 1.1e+02;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSFKPMPLXR 10

Db 236 YNLKPIPLKR 245

RESULT 5

ID Q38812 PRELIMINARY; PRT: 67 AA.
AC Q38812;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE 3-HYDROXY-3-METHYLGUTARYL COENZYME A REDUCTASE ISOFORM HMGRIL

```

db      67 YSVEPLPLFR 76
      II : I : I I I
RESULT 7
O65218 PRELIMINARY; PRT: 504 AA.
ID      O65218
AC      O65218;
DT      01-AUG-1998 (TREMBLrel. 07, Created)
DT      01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      PUTATIVE METHYLTRANSFERASE.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota: Viridiplantae: Streptophyta:
OC      Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae:
OC      eurosids II: Brassicales; Brassicaceae; Arabidopsis.
OX      NCBI_TaxID=3702;
      [1]
      RN

```

DR EMBL; AF058800; AAC14296.1; -.
DR InterPro: IP8001214; SET.

KW Transferase; Methyltransferase.
 SO SEQUENCE 504 AA; 57644 MW; B9CA9515863AAAA0 CRC64;

```

Best Local Similarity 66.7%;   Pred. No. 92;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      2 SFKPMPLXR 10
         |||.|||
Db       29 SFSPLPLSR 37

RESULT 8

```

ID	Q957D2	PRELIMINARY;	PRT;	504 AA.
AC	Q957D2;			
DT	01-MAY-2000	(TrEMBLrel. 13, Created)		
DT	01-MAY-2000	(TrEMBLrel. 13, Last sequence update)		
DT	01-JUN-2001	(TrEMBLrel. 17, Last annotation update)		
DE	PUTATIVE RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/GENASE SMALL SUBUNIT			
DE	N-METHYLTRANSFERASE 1.			
GN	F17A17.1	NP_17.12.		
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta, eudicotyledons; Core eudicots; Rosidae;			

OX	NCBI_TaxID	NCBI_TaxName
RN	11	SEQUENCE FROM N.A.
RP		SEQUENCE FROM N.A.
RC		STRAIN=CV. COLUMBIA;

RA ROBINSON C.M., KOO H., FUJII C.Y., UTTERBACK T.R., BARNSTEAD M.E.,
 RA BOWMAN C.L., WHITE O., NIERMAN W.C., PRASER C.M.,
 RA "Arabidopsis thaliana chromosome III WAC F1747 genomic sequence";
 RT Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RL
 RN [2]
 RP
 RC SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA LIN X., KAUL S., TOMI C.D., BENITO M., CREASY T.H., HAAS B.,
 RA RONNING C.M., KOO H., FUJII C.Y., UTTERBACK T.R., BARNSTEAD M.E.,
 RA

RL Submitted (NOV-1999) to the EMBL/Genbank/DDBJ databases.

DR EMBL; AC013483; AAF21177.J; -

DR EMBL; AC009176; AAF13085.1; -

DR InterPro: IPR001214; SET.

DR PROSITE; PS0280; SET; 1.

```

DR SMART: SM00317: SET: 1.
KW Transferase: Methyltransferase.
SQ SEQUENCE 504 AA; 57613 MW; A5BB41F36FAB08BE CRC64;

Query Match 68.6%; Score 35; DB 10; Length 504;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 SFKPMPLXR 10
Db 29 SFSPPLSR 37

RESULT 9
O16849 ID O16849 PRELIMINARY; PRT; 508 AA.
AC O16849
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE FORK HEAD-RELATED TRANSCRIPTION FACTOR DAF-16A2.
GN DAF-16.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX Ogg S., Paradis S., Gottlieb S., Patterson G.I., Lee L.,
RA Tissenbaum H.A., Ruvkun G.;
RT "The Fork head transcription factor DAF-16 transduces insulin-like
metabolic and longevity signals in C. elegans.";
RL Nature 389:994-999(1997).
DR EMBL: AF020343; AAB84391.1; -.
DR TRANSFAC: T03399; -.
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
SQ SEQUENCE 508 AA; 55577 MW; 2C0CF97657CD6350 CRC64;

Query Match 68.6%; Score 35; DB 5; Length 508;
Best Local Similarity 85.7%; Pred. No. 93;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
Db 415 NFKPMPL 421

RESULT 10
O18676 ID O18676 PRELIMINARY; PRT; 510 AA.
AC O18676;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE DAF-16.
GN DAF-16.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=DRISTOL N2;
RX Ogg S., Paradis S., Gottlieb S., Patterson G.I., Lee L.,
RA Tissenbaum H.A., Ruvkun G.;
RT "The Fork head transcription factor DAF-16 transduces insulin-like
metabolic and longevity signals in C. elegans.";
RL Nature 389:994-999(1997).
DR EMBL: AF020344; AAB84392.1; -.
DR HSSP: Q63245; 2HFH.
DR TRANSFAC: T03400; -.
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
SQ SEQUENCE 530 AA; 57898 MW; 8665FBD8428039D6 CRC64;

Query Match 68.6%; Score 35; DB 5; Length 530;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
Db 417 NFKPMPL 423

RESULT 11
O18850 ID O18850 PRELIMINARY; PRT; 530 AA.
AC O18850;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE FORK HEAD-RELATED TRANSCRIPTION FACTOR DAF-16B.
GN DAF-16.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX Ogg S., Paradis S., Gottlieb S., Patterson G.I., Lee L.,
RA Tissenbaum H.A., Ruvkun G.;
RT "The Fork head transcription factor DAF-16 transduces insulin-like
metabolic and longevity signals in C. elegans.";
RL Nature 389:994-999(1997).
DR EMBL: AF020342; AAB84390.1; -.
DR TRANSFAC: T03398; -.
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
SQ SEQUENCE 510 AA; 55867 MW; 52574F3F979B3583 CRC64;

Query Match 68.6%; Score 35; DB 5; Length 510;
Best Local Similarity 85.7%; Pred. No. 93;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
Db 417 NFKPMPL 423

RESULT 11
O18850 ID O18850 PRELIMINARY; PRT; 530 AA.
AC O18850;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE FORK HEAD-RELATED TRANSCRIPTION FACTOR DAF-16B.
GN DAF-16.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX Ogg S., Paradis S., Gottlieb S., Patterson G.I., Lee L.,
RA Tissenbaum H.A., Ruvkun G.;
RT "The Fork head transcription factor DAF-16 transduces insulin-like
metabolic and longevity signals in C. elegans.";
RL Nature 389:994-999(1997).
DR EMBL: AF020344; AAB84392.1; -.
DR HSSP: Q63245; 2HFH.
DR TRANSFAC: T03400; -.
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head; 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; FH; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
SQ SEQUENCE 530 AA; 57898 MW; 8665FBD8428039D6 CRC64;

Query Match 68.6%; Score 35; DB 5; Length 530;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 SFKPMPL 8
Db 417 NFKPMPL 423

```

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Basham V.; (NOV-1996) to the EMBL/GenBank/DBJ databases.
 KL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.*;
 RL Nature 368:32-38(1994).
 DR EMBL: 282278; CAB05258.1;
 DR InterPro: IPR003125; WSN.
 DR Pfam: PF02206; WSN; 1.
 DR SMART: SM00453; WSN; 1.
 DR SEQUENCE 690 AA; 78360 MW; 0E8A2B2A2547F124 CRC64;
 SQ
 Query Match 68.6%; Score 35; DB 5; Length 690;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 SFKPMPL 8
 Db 323 SFKPLPL 329
 RESULT 14
 O45958 PRELIMINARY; PRT; 717 AA.
 ID O45958
 AC O45958;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE Y51A28.6A PROTEIN.
 GN Y51A28.6A.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA McMurray A.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.*;
 RL Nature 368:32-38(1994).

Db 437 NFKPMPL 443
 RESULT 12
 O44756 PRELIMINARY; PRT; 622 AA.
 ID O44756;
 AC O44756;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE R13H8.1; PROTEIN.
 GN R13H8.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=BRISTOL N2;
 RC MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaison N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.*;
 RL Nature 368:32-38(1994).
 DR EMBL: 282278; CAB05258.1;
 DR InterPro: IPR003125; WSN.
 DR Pfam: PF02206; WSN; 1.
 DR SMART: SM00453; WSN; 1.
 DR SEQUENCE 690 AA; 78360 MW; 0E8A2B2A2547F124 CRC64;
 SQ
 Query Match 68.6%; Score 35; DB 5; Length 622;
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 SFKPMPL 8
 Db 454 NFKPMPL 460
 RESULT 13
 O9XUC6 PRELIMINARY; PRT; 690 AA.
 ID O9XUC6
 AC O9XUC6;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE M162.7 PROTEIN.
 GN M162.7.

DR EMBL: AL021493; CAA16394.1; --
 DR InterPro: IPR003125; WSN.
 DR Pfam: PF02206; WSN; 1.
 DR SMART: SM00453; WSN; 1.
 SQ SEQUENCE 717 AA; 81475 MW; EA9C79E40E7FBD6E CRC64;

Query Match 68.6%; Score 35; DB 5; Length 717;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 SFKPMPL 8
 |||||
 Db 309 SFKPLPL 315

RESULT 15
 O45959 PRELIMINARY; PRT: 717 AA.
 AC O45959;
 DT 01-JUN-1998 (TReMBLrel. 06, Created)
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE Y51A2B.6B PROTEIN.
 GN Y51A2B.6B.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA McMurray A.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Lalster N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: AL021493; CAA16395.1; --
 DR InterPro: IPR003125; WSN.
 DR Pfam: PF02206; WSN; 1.
 DR SMART: SM00453; WSN; 1.
 SQ SEQUENCE 717 AA; 81489 MW; 14E394886B373AB CRC64;

Query Match 68.6%; Score 35; DB 5; Length 717;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 SFKPMPL 8
 |||||
 Db 309 SFKPLPL 315

Search completed: February 27, 2002, 11:50:04
 Job time: 983 sec

Matches 8: Conservative 0: Mismatches 2: Indels 0: Gaps 0;

Qy 1 YSFKPMPLXR 10
|||||

Db 1 YSFKPMPLAR 10

RESULT 2

US-08-299-285-10

: Sequence 10, Application US/08299285

: Patent No. 5696230.

: GENERAL INFORMATION:

: APPLICANT: Sanderson, Sam D.

: APPLICANT: Sherman, Simon A.

: APPLICANT: Kirnarsky, Leonid

: APPLICANT: Taylor, Stephen M.

: TITLE OF INVENTION: High Affinity Response-Selective

: C-Terminal Analogs of C5a Anaphylatoxin

: NUMBER OF SEQUENCES: 37

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Dann, Dorfman, Herrell and Skillman

: STREET: 1601 Market Street Suite 720

: CITY: Philadelphia

: STATE: PA

: COUNTRY: US

: ZIP: 19103-2307

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/299,285

: FILING DATE: 31-AUG-1994

: CLASSIFICATION: 530

: ATTORNEY/AGENT INFORMATION:

: NAME: Reed, Janet E.

: REGISTRATION NUMBER: 36,252

: REFERENCE/DOCKET NUMBER: 63075

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (215) 563-4100

: TELEFAX: (215) 563-4044

: INFORMATION FOR SEQ ID NO: 10:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 10 amino acids

: TYPE: amino acid

: STRANDEDNESS: single

: TOPOLOGY: not relevant

: MOLECULE TYPE: peptide

: HYPOTHETICAL: NO

: ANTI-SENSE: NO

: FRAGMENT TYPE: C-terminal

: FEATURE:

: NAME/KEY: Modified-site

: LOCATION: 9

: OTHER INFORMATION: "D-Alanine at position 9"

US-08-299-285-10

Query Match

Best Local Similarity 82.4%; Score 42; DB 1; Length 10;

Matches 8: Conservative 0: Mismatches 2: Indels 0: Gaps 0;

Qy 1 YSFKPMPLXR 10

|||||

Db 1 YSFKPMPLAR 10

RESULT 3

US-08-299-285-24

: Sequence 24, Application US/08299285

: Patent No. 5696230

: GENERAL INFORMATION:

: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Dann, Dorfman, Herrell and Skillman

: STREET: 1601 Market Street Suite 720

: CITY: Philadelphia

: STATE: PA

: COUNTRY: US

: ZIP: 19103-2307

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/299,285

: FILING DATE: 31-AUG-1994

: CLASSIFICATION: 530

: ATTORNEY/AGENT INFORMATION:

: NAME: Reed, Janet E.

: REGISTRATION NUMBER: 36,252

: REFERENCE/DOCKET NUMBER: 63075

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (215) 563-4100

: TELEFAX: (215) 563-4044

: INFORMATION FOR SEQ ID NO: 24:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 10 amino acids

: TYPE: amino acid

: STRANDEDNESS: not relevant

: TOPOLOGY: not relevant

: MOLECULE TYPE: peptide

: HYPOTHETICAL: NO

: ANTI-SENSE: NO

: FRAGMENT TYPE: C-terminal

: FEATURE:

: NAME/KEY: Modified-site

: LOCATION: 9

: OTHER INFORMATION: "D-Alanine at position 9"

US-08-299-285-24

Query Match

Best Local Similarity 82.4%; Score 42; DB 1; Length 10;

Matches 8: Conservative 0: Mismatches 2: Indels 0: Gaps 0;

Qy 1 YSFKPMPLXR 10

|||||

Db 1 YSFKPMPLAR 10

RESULT 4

US-08-985-126-7

: Sequence 7, Application US/08985126

: Patent No. 5942599

: GENERAL INFORMATION:

: APPLICANT: Sanderson, Sam D.

: APPLICANT: Sherman, Simon A.

: APPLICANT: Kirnarsky, Leonid

: APPLICANT: Taylor, Stephen M.

: TITLE OF INVENTION: High Affinity Response-Selective

: C-Terminal Analogs of C5a Anaphylatoxin

: NUMBER OF SEQUENCES: 37

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Dann, Dorfman, Herrell and Skillman

: STREET: 1601 Market Street Suite 720

: CITY: Philadelphia

: STATE: PA

```
;
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US/08/299,285
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. 5942599 Relevant
; TOPOLOGY: No. 5942599 Relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; US-08-985-126-7
;
; Query Match 82.4%; Score 42; DB 2; Length 10;
; Best Local Similarity 80.0%; Pred. No. 0.026;
; Matches 8; Conservative 0; Mismatches 2; Indels 0; Caps 0;
;
QY 1 YSPKPMPLXR 10
DB 1 YSPKPMPLAR 10
;
; RESULT 5
; US-08-985-126-10
; Sequence 10, Application US/08985126
; Patent No. 5942599
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
```

```
;
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: No. 5942599 Relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: "D-Alanine at position 9"
; US-08-985-126-10
;
; Query Match 82.4%; Score 42; DB 2; Length 10;
; Best Local Similarity 80.0%; Pred. No. 0.026;
; Matches 8; Conservative 0; Mismatches 2; Indels 0; Caps 0;
;
QY 1 YSPKPMPLXR 10
DB 1 YSPKPMPLAR 10
;
; RESULT 6
; US-08-985-126-24
; Sequence 24, Application US/08985126
; Patent No. 5942599
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 24:
```

```
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: No. 5942599 Relevant
: TOPOLOGY: No. 5942599 Relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: "D-Alanine at position 9"
US-08-985-126-24

Query Match      82.4%; Score 42; DB 2; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.026;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSEKPMPLXR 10
Db 1 YSEKPMQLAR 10

RESULT 7
PCT-US95-11126-7
: Sequence 7, Application PC/TUS9511126
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/299,285
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: "D-Alanine at position 9"
PCT-US95-11126-7

Query Match      82.4%; Score 42; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.026;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSEKPMPLXR 10
Db 1 YSEKPMPLAR 10
```

```
Query Match      82.4%; Score 42; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.026;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSEKPMPLXR 10
Db 1 YSEKPMPLAR 10

RESULT 8
PCT-US95-11126-10
: Sequence 10, Application PC/TUS9511126
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/299,285
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: "D-Alanine at position 9"
PCT-US95-11126-10

Query Match      82.4%; Score 42; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.026;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSEKPMPLXR 10
Db 1 YSEKPMPLAR 10
```

```

RESULT 9
PCT-US95-11126-24
: Sequence 24, Application PC/TUS9511126
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/299,285
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: /product="D-Alanine at position 9"
PCT-US95-11126-24

Query Match 82.4%; Score 42; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.026;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSFKPMPLXR 10
   | | | | | | |
Db 1 YSFKPMQLAR 10

RESULT 10
US-08-299-285-5
: Sequence 5, Application US/08299285
: Patent No. 5696230
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/299,285
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 24:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: /product="D-Alanine at position 9"
PCT-US95-11126-24

```

```

: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,285
: FILING DATE: 31-AUG-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: US-08-299-285-5

Query Match 80.4%; Score 41; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.04;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSFKPMPLXR 10
   | | | | | | |
Db 1 YSFKDMPLGR 10

RESULT 11
US-08-299-285-13
: Sequence 13, Application US/08299285
: Patent No. 5696230
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: "D-Phenylalanine at
position 9"
US-08-299-285-13

Query Match 80.4%; Score 41; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.04;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YSFKPMPLXR 10
Db 1 YSFKDMPPLR 10

RESULT 12
US-08-985-126-5
Sequence 5, Application US/08985126
Patent No. 5942599
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:

TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: No. 5942599 Relevant
TOPOLOGY: No. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-985-126-5

Query Match 80.4%; Score 41; DB 2; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.04;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 YSFKPMPLXR 10
Db 1 YSFKDMPPLR 10

RESULT 13
US-08-985-126-13
Sequence 13, Application US/08985126
Patent No. 5942599
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: No. 5942599 Relevant
TOPOLOGY: No. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:

```
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: "D-Phenylalanine at
: OTHER INFORMATION: position 9"
US-08-985-126-13

Query Match      80.4%; Score 41; DB 2; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.04;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSRKPMPLXR 10
   111111111
Db 1 YSKDMPFLR 10

RESULT 14
PCT-US95-11126-5
: Sequence 5, Application PC/TUS9511126
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnatsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: TITLE OF INVENTION: C-terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA: US 08/299,285
: APPLICATION NUMBER: 31-AUG-1994
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: ORIGINAL SOURCE:
: ORGANISM: C5a Anaphylatoxin
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: /product="D-Phenylalanine at
: OTHER INFORMATION: position 9"
PCT-US95-11126-5

Query Match      80.4%; Score 41; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.04;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSRKPMPLXR 10
   111111111
Db 1 YSKDMPFLR 10

RESULT 15
PCT-US95-11126-13
: Sequence 13, Application PC/TUS9511126
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnatsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: TITLE OF INVENTION: C-terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA: US 08/299,285
: APPLICATION NUMBER: 31-AUG-1994
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: ORIGINAL SOURCE:
: ORGANISM: C5a Anaphylatoxin
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: /product="D-Phenylalanine at
: OTHER INFORMATION: position 9"
PCT-US95-11126-13

Query Match      80.4%; Score 41; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.04;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 YSRKPMPLXR 10
   111111111
Db 1 YSKDMPFLR 10

Search completed: February 27, 2002, 11:36:02
Job time: 141 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:13 ; Search time 303.5 Seconds
(without alignments)
5.125 Million cell updates/sec

Title: US-09-446-109A-5
Perfect score: 100
Sequence: 1 RAARISLGRXYSFKPMPLXR 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. Is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51	51.0	12	AAW27254	Molecular adjuvant
2	50	50.0	10	AAW27250	Molecular adjuvant
3	50	50.0	18	AAW27256	Molecular adjuvant
4	50	50.0	19	AAW27251	Molecular adjuvant
5	48	48.0	64	AAW07784	Human complement C
6	48	48.0	65	AAW07785	Human complement C
7	48	48.0	66	AAW07786	Human complement C
8	48	48.0	67	AAW07787	Human complement C
9	48	48.0	68	AAW07788	Human complement C
10	48	48.0	69	AAW07789	Human complement C
11	48	48.0	70	AAW07790	Human complement C

12	48	48.0	71	18	AAW07782	Human C5a
13	48	48.0	71	18	AAW07783	Human complement C
14	48	48.0	71	18	AAW07804	Human complement C
15	48	48.0	71	18	AAW07791	Human complement C
16	48	48.0	72	18	AAW07792	Human complement C
17	48	48.0	74	8	AAW71666	Human anaphylatoxin
18	48	48.0	74	16	AAW75497	Human C5a protein
19	48	48.0	74	22	AAE05454	Human C5a anaphyla
20	48	48.0	74	22	AAE05453	Human C5a. Homo s
21	48	48.0	74	22	AAW74119	Variant human C5a.
22	48	48.0	1676	16	AAW77604	Pro-C5 polypeptide
23	45	45.0	74	20	AAW95580	Solid phase sequen
24	43	43.0	9	20	AAW74061	Human C5a peptide
25	43	43.0	758	20	AAW35578	C. pneumoniae prot
26	42	42.0	10	17	AAW94487	C5a anaphylatoxin
27	42	42.0	10	17	AAW94470	C5a anaphylatoxin
28	42	42.0	10	17	AAW94473	C5a anaphylatoxin
29	42	42.0	263	21	AAW27308	Arabidopsis thalia
30	42	42.0	280	21	AAW27307	Arabidopsis thalia
31	42	42.0	285	21	AAW27306	Arabidopsis thalia
32	41	41.0	9	22	AAW74071	C-terminal truncat
33	41	41.0	10	17	AAW94476	C5a anaphylatoxin
34	41	41.0	10	17	AAW94477	C5a anaphylatoxin
35	41	41.0	117	22	AAW01570	Human secreted pro
36	40	40.0	9	22	AAW74072	C-terminal truncat
37	40	40.0	9	22	AAW74073	C-terminal truncat
38	40	40.0	10	17	AAW94486	C5a anaphylatoxin
39	40	40.0	10	17	AAW94477	C5a anaphylatoxin
40	40	40.0	77	22	AAE05458	Mouse C5a anaphyla
41	40	40.0	180	22	AAW47008	L. intracellularis
42	40	40.0	381	22	AAW52461	Mycobacterium tube
43	40	40.0	429	22	AAW6342	H. pylori HPS117 p
44	40	40.0	429	22	AAW6403	H. pylori HPS117 p
45	39.5	39.5	104	22	AAW95652	Human protein sequ

ALIGNMENTS

RESULT	1				
AAW27254					
ID	AAW27254	standard; peptide; 12 AA.			
XX	AAW27254:				
AC					
DT	23-DEC-1997	(first entry)			
XX					
DE		Molecular adjuvant formula.			
XX					
KW		Molecular adjuvant; immune response; immunogen; binding affinity;			
KW		antigen presenting cell; APC; viral pathogen; anti tumour response;			
KW		antibody; immunodiagnostic; immunotherapeutic; agonist.			
XX					
OS		Synthetic.			
XX					
FT	Key	Location/Qualifiers			
FT	Modified-site 1	/note= "Serum amyloid A is attached to the N-terminal"			
FT	Modified-site 2	/label= OTHER			
FT		/note= "Aminohexanoic acid which is a linear aliphatic spacer moiety"			
FT	Misc-difference 11	/note= "D-form residue"			
FT					
XX					
PN	WO9714426-A1.				
XX					
PD	24-APR-1997.				
XX					
PF	18-OCT-1996;	96WO-US16825.			
XX					
PR	20-OCT-1995;	95US-0005727.			
XX					

PA (UYNE-) UNIV NEBRASKA.
 XX Hollingsworth MA, Sanderson SD, Tempero RA;
 PT WPI: 1997-244854/22.
 XX Molecular adjuvants for enhancing immune responses - comprise an
 XX immunogen linked to a ligand having binding affinity for a
 PT determinant of an antigen presenting cell
 PT
 XX Claim 15; Page 53; 6lpp; English.
 XX A novel molecular adjuvant has been developed for enhancing an immune
 CC response to an immunogen. The molecular adjuvant comprises a targeting
 CC ligand, having binding affinity for a characteristic determinant of an
 CC antigen presenting cell (APC), the targeting ligand being functionally
 CC linked to the immunogen, and so binding of the molecular adjuvant to
 CC the APC determinant activates the APC, effecting delivery of the
 CC immunogen to an antigen presenting pathway of the APC. The present
 CC sequence represents a specifically claimed formula for a molecular
 CC adjuvant, where the immunogen is serum amyloid A. The molecular adjuvant
 CC can be used to produce an immune response for protecting against viral
 CC and other pathogens or to produce anti-tumour responses. It can also be
 CC used for the production of antibodies for use as immunodiagnostic and
 CC immunotherapeutic agents. The molecular adjuvant is readily taken up and
 CC processed by APCs to provide APC-mediated immune responses. It can
 CC provide for the production of antibodies against a weakly-antigenic or
 CC non-antigenic substances.
 XX Sequence 12 AA;

Query Match 51.0%; Score 51; DB 18; Length 12;
 Best Local Similarity 83.3%; Pred. No. 0.054;
 Matches 10; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 10 RXYSFKPMPLXR 21
 :|||||||
 Db 1 kxyfkmplrar 12

.RESULT 2
 AAW27250
 ID AAW27250 standard; peptide; 10 AA.
 XX
 AC AAW27250:

XX 23-DEC-1997 (first entry)
 XX Molecular adjuvant targeting ligand.
 DE
 XX Molecular adjuvant; immune response; immunogen; binding affinity;
 KW antigen presenting cell; APC; viral pathogen; anti tumour response;
 KW antibody; immunodiagnostic; immunotherapeutic; agonist.
 XX
 OS Synthetic.

XX Key Location/Qualifiers
 FH Misc-difference 9 /note= "D-form residue"
 FT
 FT

XX W09714426-A1.
 XX 24-APR-1997.
 XX 18-OCT-1996; 96WO-US16825.
 XX 20-OCT-1995; 95US-0005727.
 XX (UYNE-) UNIV NEBRASKA.

XX Hollingsworth MA, Sanderson SD, Tempero RA;
 PT WPI: 1997-244854/22.
 XX Molecular adjuvants for enhancing immune responses - comprise an
 XX immunogen linked to a ligand having binding affinity for a

DR WPI: 1997-244854/22.
 XX Molecular adjuvants for enhancing immune responses - comprise an
 PT immunogen linked to a ligand having binding affinity for a
 PT determinant of an antigen presenting cell
 XX
 PS Claim 5; Page 51; 6lpp; English.
 XX A novel molecular adjuvant has been developed for enhancing an immune
 CC response to an immunogen. The molecular adjuvant comprises a targeting
 CC ligand, having binding affinity for a characteristic determinant of an
 CC antigen presenting cell (APC), the targeting ligand being functionally
 CC linked to the immunogen, and so binding of the molecular adjuvant to
 CC the APC determinant activates the APC, effecting delivery of the
 CC immunogen to an antigen presenting pathway of the APC. The present
 CC sequence represents a specifically claimed targeting ligand, which is
 CC a C α -terminal decapeptide agonist. The molecular adjuvant can be used
 CC to produce an immune response for protecting against viral and other
 CC pathogens or to produce anti-tumour responses. It can also be used for
 CC the production of antibodies for use as immunodiagnostic and
 CC immunotherapeutic agents. The molecular adjuvant is readily taken up and
 CC processed by APCs to provide APC-mediated immune responses. It can
 CC provide for the production of antibodies against a weakly-antigenic or
 CC non-antigenic substances.
 XX Sequence 10 AA;

Query Match 50.0%; Score 50; DB 18; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.066;
 Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 YSEKPMPLXR 21
 :|||||||
 Db 1 ysfkpmplrar 10

RESULT 3
 AAW27256
 ID AAW27256 standard; peptide; 18 AA.
 XX
 AC AAW27256:

XX 23-DEC-1997 (first entry)

XX Molecular adjuvant targeting ligand and immunogen.

XX Molecular adjuvant; immune response; immunogen; binding affinity;
 KW antigen presenting cell; APC; viral pathogen; anti tumour response;
 KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX Synthetic.

XX Key Location/Qualifiers
 FH Misc-difference 9 /note= "D-form residue"
 FT

XX W09714426-A1.
 XX 24-APR-1997.
 XX 18-OCT-1996; 96WO-US16825.
 XX 20-OCT-1995; 95US-0005727.
 XX (UYNE-) UNIV NEBRASKA.

XX Hollingsworth MA, Sanderson SD, Tempero RA;
 PT WPI: 1997-244854/22.
 XX Molecular adjuvants for enhancing immune responses - comprise an
 XX immunogen linked to a ligand having binding affinity for a

determinant of an antigen presenting cell

Example 1; Page 31; 61pp; English.

A novel molecular adjuvant has been developed for enhancing an immune response to an immunogen. The molecular adjuvant comprises a targeting ligand, having binding affinity for a characteristic determinant of an antigen presenting cell (APC), the targeting ligand being functionally linked to the immunogen, and so binding of the molecular adjuvant to the APC determinant activates the APC, effecting delivery of the immunogen to an antigen presenting pathway of the APC. The present sequence represents a targeting ligand and immunogen, in which the juxta-membrane epitope of human mucin-1 is positioned toward the carboxyl terminus and the C5a analogue is positioned toward the amino terminus. The molecular adjuvant can be used to produce an immune response for protecting against viral and other pathogens or to produce anti-tumour responses. It can also be used for the production of antibodies for use as immunodiagnostic and immunotherapeutic agents. The molecular adjuvant is readily taken up and processed by APCs to provide APC-mediated immune responses. It can provide for the production of antibodies against a weakly-antigenic or non-antigenic substances.

Sequence 18 AA;

Query Match 50.0%; Score 50; DB 18; Length 18;

Best Local Similarity 90.0%; Pred. No. 0.12;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 YSPKPMPLXR 21

Db 1 ysfkpmplar 10

RESULT 4

AAW27251

ID AAW27251 standard; peptide; 19 AA.

XX XX

AC AAW27251;

XX 23-DEC-1997 (first entry)

XX Molecular adjuvant targeting ligand and immunogen.

XX Molecular adjuvant; immune response; immunogen; binding affinity;
KW antigen presenting cell; APC; viral pathogen; anti tumour response;
KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 18 /note= "D-form residue"

XX W09714426-A1.

PN 24-APR-1997.

XX 18-OCT-1996; 96WO-US16825.

PF 20-OCT-1995; 95US-0005727.

XX (UYNE-) UNIV NEBRASKA.

XX Hollingsworth MA, Sanderson SD, Tempero RA;

PI WPI; 1997-244854/22.

XX Molecular adjuvants for enhancing immune responses - comprise an
PT immunogen linked to a ligand having binding affinity for a
PT determinant of an antigen presenting cell

PS Claim 6; Page 52; 61pp; English.

XX A novel molecular adjuvant has been developed for enhancing an immune response to an immunogen. The molecular adjuvant comprises a targeting ligand, having binding affinity for a characteristic determinant of an antigen presenting cell (APC), the targeting ligand being functionally linked to the immunogen, and so binding of the molecular adjuvant to the APC determinant activates the APC, effecting delivery of the immunogen to an antigen presenting pathway of the APC. The present sequence represents a specifically claimed targeting ligand and immunogen. The molecular adjuvant can be used to produce an immune response for protecting against viral and other pathogens or to produce anti-tumour responses. It can also be used for the production of antibodies for use as immunodiagnostic and immunotherapeutic agents. The molecular adjuvant is readily taken up and processed by APCs to provide APC-mediated immune responses. It can provide for the production of antibodies against a weakly-antigenic or non-antigenic substances.

XX Sequence 19 AA;

Query Match 50.0%; Score 50; DB 18; Length 19;

Best Local Similarity 90.0%; Pred. No. 0.13;

Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 YSPKPMPLXR 21

Db 10 ysfkpmplar 19

RESULT 5

AAW07784

ID AAW07784 standard; protein; 64 AA.

XX XX

AC AAW07784;

XX 01-SEP-1997 (first entry)

XX Human complement C5a protein derivative analogue 1.

XX Human; complement; C5a; derivative; receptor; antagonist; trauma;
KW treatment; prevention; disease; inflammation; pneumonitis; burn;
KW adult respiratory distress syndrome; ARDS; pulmonary; injury;
KW post myocardial infarction; inflammatory bowel; endotoxic shock;
KW rheumatoid arthritis; psoriasis; sepsis; transplant rejection;
KW immunosuppressive therapy; blood transfusion; dysfunction;
KW haemodialysis; leukopheresis; prophylaxis; reperfusion.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 1 /note= "wild type Thr replaced by Met"

FT Misc-difference 27 /note= "wild type Cys replaced by Ser"

FT Misc-difference 64 /note= "wild type Asn replaced by Cys"

XX W09639503-A1.

XX 12-DEC-1996.

XX 04-JUN-1996; 96WO-EP02422.

XX 05-JUN-1995; 95US-0463377.

XX 05-JUN-1995; 95US-0462648.

XX 05-JUN-1995; 95US-0463224.

XX (CIBA) CIBA GEIGY AG.

XX Schmitz A, Van Heeke G, Van Oostrum J;

XX WPI: 1997-043125/04.
 XX
 PT New human complement C5a poly-peptide derivs. - used as C5a receptor
 PT antagonists, partic. for treating C5a-mediated diseases and
 PT inflammatory conditions
 XX
 PS Example 3; Page -: 93pp; English.
 XX
 CC The present sequence, the human complement C5a derivative 1-64,
 CC ThrMet, Cys27Ser, Asn64Cys, is a C5a receptor antagonist which
 CC exhibits no agonist activity. It can be used to treat or prevent
 CC C5a mediated diseases or inflammation, e.g. pneumonitis, adult
 CC respiratory distress syndrome (ARDS), pulmonary inflammation or
 CC injury, post myocardial infarction inflammation, inflammatory bowel
 CC disease, rheumatoid arthritis, psoriasis, endotoxemic shock, sepsis,
 CC severe trauma and burns. It can also be used to treat patients
 CC suffering from transplant rejection, receiving immunosuppressive
 CC therapy or massive blood transfusion, exposed to medical devices
 CC or experiencing pulmonary dysfunction following haemodialysis or
 CC leukopheresis. It can also be used as a prophylactic, particularly
 CC in conditions caused by reperfusion, e.g. reperfusion following
 CC ischaemia, and circulatory contact with medical devices, as well as
 CC to prevent transplant rejection.
 CC Antibodies against the derivative can be used to detect or quantify
 CC the derivative and modify, e.g. neutralise, its activity in vivo.
 CC N.B. Sequence not given in specification, but constructed using the
 CC wild type sequence given on pages 51-52.
 XX
 SQ Sequence 64 AA:
 Query Match 48.0%; Score 48; DB 18; Length 64;
 Best Local Similarity 100.0%; Pred. No. 0.95;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0:
 OY 1 RAARISLGR 10
 Db | | | | | | | | | |
 37 raarislgpr 46
 RESULT 6
 AAW07785
 ID AAW07785 standard; protein; 65 AA.
 XX
 AC AAW07785;
 XX
 DT 01-SRP-1997 (first entry)
 XX
 DE Human complement C5a protein derivative analogue 2.
 XX
 KW Human; complement: C5a; derivative; receptor; antagonist; trauma;
 KW treatment; prevention; disease; inflammation; pneumonitis; burn;
 KW adult respiratory distress syndrome; ARDS; pulmonary; injury;
 KW post myocardial; infarction; inflammatory bowel; endotoxemic shock;
 KW rheumatoid arthritis; psoriasis; sepsis; transplant rejection;
 KW immunosuppressive therapy; blood transfusion; dysfunction;
 KW haemodialysis; leukopheresis; prophylaxis; reperfusion.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FI Key Location/Qualifiers
 FT Misc-difference 1 /note- "wild type Thr replaced by Met"
 FT Misc-difference 27 /note- "wild type Cys replaced by Ser"
 FT Misc-difference 65 /note- "wild type Ile replaced by Cys"
 FT
 XX WO9639503-A1.
 XX
 XX 12-DKC-1996.
 PD

XX 04-JUN-1996; 95WO-EP02422.
 XX
 PR 05-JUN-1995; 95US-0463377.
 PR 05-JUN-1995; 95US-0462648.
 PR 05-JUN-1995; 95US-0463224.
 XX
 PA (CIBA) CIBA GEIGY AG.
 XX
 PI Schmitz A, Van Heeke G, Van Oostrum J;
 XX WPI: 1997-043125/04.
 XX
 PT New human complement C5a poly-peptide derivs. - used as C5a receptor
 PT antagonists, partic. for treating C5a-mediated diseases and
 PT inflammatory conditions
 XX
 PS Example 3; Page -: 93pp; English.
 XX
 CC The present sequence, the human complement C5a derivative 1-65,
 CC ThrMet, Cys27Ser, Ile65Cys, is a C5a receptor antagonist which
 CC exhibits no agonist activity. It can be used to treat or prevent
 CC C5a mediated diseases or inflammation, e.g. pneumonitis, adult
 CC respiratory distress syndrome (ARDS), pulmonary inflammation or
 CC injury, post myocardial infarction inflammation, inflammatory bowel
 CC disease, rheumatoid arthritis, psoriasis, endotoxemic shock, sepsis,
 CC severe trauma and burns. It can also be used to treat patients
 CC suffering from transplant rejection, receiving immunosuppressive
 CC therapy or massive blood transfusion, exposed to medical devices
 CC or experiencing pulmonary dysfunction following haemodialysis or
 CC leukopheresis. It can also be used as a prophylactic, particularly
 CC in conditions caused by reperfusion, e.g. reperfusion following
 CC ischaemia, and circulatory contact with medical devices, as well as
 CC to prevent transplant rejection.
 CC Antibodies against the derivative can be used to detect or quantify
 CC the derivative and modify, e.g. neutralise, its activity in vivo.
 CC N.B. Sequence not given in specification, but constructed using the
 CC wild type sequence given on pages 51-52.
 XX
 SQ Sequence 65 AA:
 Query Match 48.0%; Score 48; DB 18; Length 65;
 Best Local Similarity 100.0%; Pred. No. 0.96;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0:
 OY 1 RAARISLGR 10
 Db | | | | | | | | | |
 37 raarislgpr 46
 RESULT 7
 AAW07786
 ID AAW07786 standard; protein; 66 AA.
 XX
 AC AAW07786;
 XX
 DT 01-SEP-1997 (first entry)
 XX
 DE Human complement C5a protein derivative analogue 3.
 XX
 KW Human; complement: C5a; derivative; receptor; antagonist; trauma;
 KW treatment; prevention; disease; inflammation; pneumonitis; burn;
 KW adult respiratory distress syndrome; ARDS; pulmonary; injury;
 KW post myocardial; infarction; inflammatory bowel; endotoxemic shock;
 KW rheumatoid arthritis; psoriasis; sepsis; transplant rejection;
 KW immunosuppressive therapy; blood transfusion; dysfunction;
 KW haemodialysis; leukopheresis; prophylaxis; reperfusion.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FI Key Location/Qualifiers
 FT Misc-difference 1 /note- "wild type Thr replaced by Met"
 FT Misc-difference 27 /note- "wild type Cys replaced by Ser"
 FT Misc-difference 65 /note- "wild type Ile replaced by Cys"
 FT
 XX WO9639503-A1.
 XX
 XX 12-DKC-1996.
 PD

```

FT Misc-difference 1 /note= "wild type Thr replaced by Met"
FT Misc-difference 27 /note= "wild type Cys replaced by Ser"
FT Misc-difference 66 /note= "wild type Ser replaced by Cys"
XX
XX WO9639503-A1.
XX
XX 12-DEC-1996.
XX
XX 04-JUN-1996; 96WO-EP02422.
XX
XX 05-JUN-1995; 95US-0463377.
XX 05-JUN-1995; 95US-0462648.
XX 05-JUN-1995; 95US-0463224.
XX
XX (CIBA ) CIBA GEIGY AG.
XX
XX Schmitz A, Van Heeke G, Van Oostrum J;
XX WPI: 1997-043125/04.
XX
XX New human complement C5a poly-peptide derivs. - used as C5a receptor
XX antagonists, partic. for treating C5a-mediated diseases and
XX inflammatory conditions
XX
XX Example 3; Page -: 93pp; English.
XX
XX The present sequence, the human complement C5a derivative 1-66,
XX ThrMet, Cys27Ser, Ser66Cys, is a C5a receptor antagonist which
XX exhibits no agonist activity. It can be used to treat or prevent
XX C5a mediated diseases or inflammation, e.g. pneumonitis, adult
XX respiratory distress syndrome (ARDS), pulmonary inflammation or
XX injury, post myocardial infarction inflammation, inflammatory bowel
XX disease, rheumatoid arthritis, psoriasis, endotoxemic shock, sepsis,
XX severe trauma and burns. It can also be used to treat patients
XX suffering from transplant rejection, receiving immunosuppressive
XX therapy or massive blood transfusion, exposed to medical devices
XX or experiencing pulmonary dysfunction following haemodialysis or
XX leukopheresis. It can also be used as a prophylactic, particularly
XX in conditions caused by reperfusion, e.g. reperfusion following
XX ischaemia, and circulatory contact with medical devices, as well as
XX to prevent transplant rejection.
XX Antibodies against the derivative can be used to detect or quantify
XX the derivative and modify, e.g. neutralise, its activity in vivo.
XX N.B. Sequence not given in specification, but constructed using the
XX wild type sequence given on pages 51-52.
XX
XX Sequence 66 AA:
XX
XX Query Match 48.0%; Score 48; DB 18; Length 66;
XX Best Local Similarity 100.0%; Pred. No. 0.98;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
XX
XX QY 1 RAARISLGPR 10
XX | | | | | | | |
XX Db 37 raarislgpr 46
XX
XX RESULT 8
XX AAW07787
XX ID AAW07787 standard; protein: 67 AA.
XX AC AAW07787;
XX
XX XX 01-SEP-1997 (first entry)
XX
XX Human complement C5a protein derivative analogue 4.
XX
XX Human: complement; C5a; derivative; receptor; antagonist; trauma;
XX treatment; prevention; disease; inflammation; pneumonitis; burn;

```

```

KW adult respiratory distress syndrome; ARDS; pulmonary; injury;
KW post myocardial; infarction; inflammatory bowel; endotoxemic shock;
KW rheumatoid arthritis; psoriasis; sepsis; transplant rejection;
KW immunosuppressive therapy; blood transfusion; dysfunction;
KW haemodialysis; leukopheresis; prophylaxis; reperfusion.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Misc-difference 1 /note= "wild type Thr replaced by Met"
XX Misc-difference 27 /note= "wild type Cys replaced by Ser"
XX Misc-difference 67 /note= "wild type His replaced by Cys"
XX
XX WO9639503-A1.
XX
XX 12-DEC-1996.
XX
XX 04-JUN-1996; 96WO-EP02422.
XX
XX 05-JUN-1995; 95US-0463377.
XX 05-JUN-1995; 95US-0462648.
XX 05-JUN-1995; 95US-0463224.
XX
XX (CIBA ) CIBA GEIGY AG.
XX
XX Schmitz A, Van Heeke G, Van Oostrum J;
XX WPI: 1997-043125/04.
XX
XX New human complement C5a poly-peptide derivs. - used as C5a receptor
XX antagonists, partic. for treating C5a-mediated diseases and
XX inflammatory conditions
XX
XX Example 3; Page -: 93pp; English.
XX
XX The present sequence, the human complement C5a derivative 1-67,
XX ThrMet, Cys27Ser, His67Cys, is a C5a receptor antagonist which
XX exhibits no agonist activity. It can be used to treat or prevent
XX C5a mediated diseases or inflammation, e.g. pneumonitis, adult
XX respiratory distress syndrome (ARDS), pulmonary inflammation or
XX injury, post myocardial infarction inflammation, inflammatory bowel
XX disease, rheumatoid arthritis, psoriasis, endotoxemic shock, sepsis,
XX severe trauma and burns. It can also be used to treat patients
XX suffering from transplant rejection, receiving immunosuppressive
XX therapy or massive blood transfusion, exposed to medical devices
XX or experiencing pulmonary dysfunction following haemodialysis or
XX leukopheresis. It can also be used as a prophylactic, particularly
XX in conditions caused by reperfusion, e.g. reperfusion following
XX ischaemia, and circulatory contact with medical devices, as well as
XX to prevent transplant rejection.
XX Antibodies against the derivative can be used to detect or quantify
XX the derivative and modify, e.g. neutralise, its activity in vivo.
XX N.B. Sequence not given in specification, but constructed using the
XX wild type sequence given on pages 51-52.
XX
XX Sequence 67 AA:
XX
XX Query Match 48.0%; Score 48; DB 18; Length 67;
XX Best Local Similarity 100.0%; Pred. No. 0.99;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Caps 0;
XX
XX QY 1 RAARISLGPR 10
XX | | | | | | | |
XX Db 37 raarislgpr 46
XX
XX RESULT 9
XX AAW07788

```


CC to prevent transplant rejection.
 CC Antibodies against the derivative can be used to detect or quantify
 CC the derivative and modify, e.g. neutralise, its activity in vivo.
 CC N.B. Sequence not given in specification, but constructed using the
 CC wild type sequence given on pages 51-52.
 CC
 SQ Sequence 69 AA;

Query Match 48.0%; Score 48; DB 18; Length 69;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RAARISLGPR 10
 Db 37 raarislgpr 46

RESULT 11
 AAW07790
 ID AAW07790 standard; protein: 70 AA.
 AC AAW07790;
 DT 01-SEP-1997 (first entry)
 DE Human complement C5a protein derivative analogue 7.
 XX
 KW Human; complement; C5a; derivative; receptor; antagonist; trauma;
 KW treatment; prevention; disease; inflammation; pneumonitis; burn;
 KW adult respiratory distress syndrome; ARDS; pulmonary; injury;
 KW post myocardial infarction; inflammation; inflammatory bowel; endotoxemic shock;
 KW rheumatoid arthritis; psoriasis; sepsis; transplant rejection;
 KW immunosuppressive therapy; blood transfusion; dysfunction;
 KW haemodialysis; leukopheresis; prophylaxis; reperfusion.
 XX
 OS Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT Misc-difference 1 /note= "wild type Thr replaced by Met"
 FT Misc-difference 27 /note= "wild type Cys replaced by Ser"
 FT Misc-difference 70 /note= "wild type Met replaced by Cys"
 XX
 PN WO9639503-A1.
 XX
 PD 12-DEC-1996.
 XX
 PF 04-JUN-1996; 96WO-EP02422.
 XX
 PR 05-JUN-1995; 95US-0463377.
 PR 05-JUN-1995; 95US-0462648.
 PR 05-JUN-1995; 95US-0463224.
 XX
 PA (CIBA) CIBA GEIGY AG.
 XX
 PI Schmitz A, Van Heeke G, Van Oostrum J;
 XX
 DR WPI; 1997-043125/04.
 XX
 PT New human complement C5a poly-peptide derivs. - used as C5a receptor
 PT antagonists, partic. for treating C5a-mediated diseases and
 PT inflammatory conditions
 XX
 PS Example 3; Page -: 93pp; English.
 XX
 CC The present sequence, the human complement C5a derivative 1-70,
 CC ThrMet, Cys27Ser, Met70Cys, is a C5a receptor antagonist which
 CC exhibits no agonist activity. It can be used to treat or prevent
 CC C5a mediated diseases or inflammation, e.g. pneumonitis, adult

CC respiratory distress syndrome (ARDS), pulmonary inflammation or
 CC injury, post myocardial infarction inflammation, inflammatory bowel
 CC disease, rheumatoid arthritis, psoriasis, endotoxemic shock, sepsis,
 CC severe trauma and burns. It can also be used to treat patients
 CC suffering from transplant rejection, receiving immunosuppressive
 CC therapy or massive blood transfusion, exposed to medical devices
 CC or experiencing pulmonary dysfunction following haemodialysis or
 CC leukopheresis. It can also be used as a prophylactic, particularly
 CC in conditions caused by reperfusion, e.g. reperfusion following
 CC ischaemia, and circulatory contact with medical devices, as well as
 CC to prevent transplant rejection.
 CC Antibodies against the derivative can be used to detect or quantify
 CC the derivative and modify, e.g. neutralise, its activity in vivo.
 CC N.B. Sequence not given in specification, but constructed using the
 CC wild type sequence given on pages 51-52.
 CC
 SQ Sequence 70 AA;

Query Match 48.0%; Score 48; DB 18; Length 70;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RAARISLGPR 10
 Db 37 raarislgpr 46

RESULT 12
 AAW07782
 ID AAW07782 standard; protein: 71 AA.
 AC AAW07782;
 DT 01-SEP-1997 (first entry)
 DE Human complement C5a protein derivative.
 XX
 KW Human; complement; C5a; derivative; receptor; antagonist; trauma;
 KW treatment; prevention; disease; inflammation; pneumonitis; burn;
 KW adult respiratory distress syndrome; ARDS; pulmonary; injury;
 KW post myocardial infarction; inflammation; inflammatory bowel; endotoxemic shock;
 KW rheumatoid arthritis; psoriasis; sepsis; transplant rejection;
 KW immunosuppressive therapy; blood transfusion; dysfunction;
 KW haemodialysis; leukopheresis; prophylaxis; reperfusion.
 XX
 OS Homo sapiens.
 OS Synthetic.

XX Key Location/Qualifiers
 FT Misc-difference 1 /note= "wild type Thr replaced by Gly"
 FT Misc-difference 27 /note= "wild type Cys replaced by Ser"
 FT Misc-difference 71 /note= "wild type Cln replaced by Cys"
 XX
 PN WO9639503-A1.
 XX
 PD 12-DEC-1996.
 XX
 PF 04-JUN-1996; 96WO-EP02422.
 XX
 PR 05-JUN-1995; 95US-0463377.
 PR 05-JUN-1995; 95US-0462648.
 PR 05-JUN-1995; 95US-0463224.
 XX
 PA (CIBA) CIBA GEIGY AG.
 XX
 PI Schmitz A, Van Heeke G, Van Oostrum J;
 XX
 DR WPI; 1997-043125/04.
 XX

PT New human complement C5a poly-peptide derivs. - used as C5a receptor
PT antagonists, partic. for treating C5a-mediated diseases and
PT inflammatory conditions

XX Claim 11; Page -: 93pp; English.

XX The present sequence, the human complement C5a derivative 1-71,
CC ThrGly, Cys27Ser, Gln71Cys, is a C5a receptor antagonist which
CC exhibits no agonist activity. It can be used to treat or prevent
CC C5a mediated diseases or inflammation, e.g. pneumonitis, adult
CC respiratory distress syndrome (ARDS), pulmonary inflammation or
CC injury, post myocardial infarction inflammation, inflammatory bowel
CC disease, rheumatoid arthritis, psoriasis, endotoxemic shock, sepsis,
CC severe trauma and burns. It can also be used to treat patients
CC suffering from transplant rejection, receiving immunosuppressive
CC therapy or massive blood transfusion, exposed to medical devices
CC or experiencing pulmonary dysfunction following haemodialysis or
CC leukopheresis. It can also be used as a prophylactic, particularly
CC in conditions caused by reperfusion, e.g. reperfusion following
CC ischaemia, and circulatory contact with medical devices, as well as
CC to prevent transplant rejection.
CC Antibodies against the derivative can be used to detect or quantify
CC the derivative and modify, e.g. neutralise, its activity in vivo.
CC N.B. Sequence not given in specification, but constructed using the
CC wild type sequence given on pages 51-52.

XX Sequence 71 AA;

Query Match 48.0%; Score 48; DB 18; Length 71;
Best Local Similarity 100.0%; Pred. No. 1,1,1; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

QY 1 RAARISIGPR 10
| | | | | | | | | |
Db 37 raarisigpr 46

RESULT 13

AAW07783
ID AAW07783 standard; protein; 71 AA.

XX AAW07783;

XX 01-SEP-1997 (first entry)

XX Human complement C5a protein derivative.

XX Human; complement; C5a; derivative; receptor; antagonist; trauma;
XX treatment; prevention; disease; inflammation; pneumonitis; burn;
XX adult respiratory distress syndrome; ARDS; pulmonary; injury;
XX post myocardial; infarction; inflammatory bowel; endotoxemic shock;
XX rheumatoid arthritis; psoriasis; sepsis; transplant rejection;
XX immunosuppressive therapy; blood transfusion; dysfunction;
XX haemodialysis; leukopheresis; prophylaxis; reperfusion.

XX Homo sapiens.

XX Synthetic.

XX Key Location/Qualifiers

XX Misc-difference 1 /note= "wild type Thr replaced by Gly"

XX Misc-difference 27 /note= "wild type Cys replaced by Ser"

XX Misc-difference 67 /note= "wild type His replaced by Phe"

XX Misc-difference 71 /note= "wild type Gln replaced by Cys"

XX WO9639503-A1.

XX 12-DEC-1996.

XX

XX

XX

XX

PF 04-JUN-1996; 96WO-EP02422.

XX 05-JUN-1995; 95US-0463377.

PR 05-JUN-1995; 95US-0462648.

XX 05-JUN-1995; 95US-0463224.

XX (CIBA) CIBA GEIGY AG.

XX Schmitz A, Van Heeke G, Van Oostrum J;

XX WPI; 1997-043125/04.

XX New human complement C5a poly-peptide derivs. - used as C5a receptor

XX antagonists, partic. for treating C5a-mediated diseases and

XX inflammatory conditions

XX Claim 12; Page -: 93pp; English.

XX The present sequence, the human complement C5a derivative 1-71,

XX ThrGly, Cys27Ser, His67Phe, Gln71Cys, is a C5a receptor antagonist

XX which exhibits no agonist activity. It can be used to treat or

XX prevent C5a mediated diseases or inflammation, e.g. pneumonitis,

XX adult respiratory distress syndrome (ARDS), pulmonary inflammation

XX or injury, post myocardial infarction inflammation, inflammatory

XX bowel disease, rheumatoid arthritis, psoriasis, endotoxemic shock,

XX sepsis, severe trauma and burns. It can also be used to treat

XX patients suffering from transplant rejection, receiving

XX immunosuppressive therapy or massive blood transfusion, exposed to

XX medical devices or experiencing pulmonary dysfunction following

XX haemodialysis or leukopheresis. It can also be used as a

XX prophylactic, particularly in conditions caused by reperfusion,

XX e.g. reperfusion following ischaemia, and circulatory contact with

XX medical devices, as well as to prevent transplant rejection.

XX Antibodies against the derivative can be used to detect or quantify

XX the derivative and modify, e.g. neutralise, its activity in vivo.

XX N.B. Sequence not given in specification, but constructed using the

XX wild type sequence given on pages 51-52.

XX Sequence 71 AA;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Search completed: February 27, 2002, 11:41:14
Job time: 453 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:15 : Search time 145.23 Seconds
(without alignments)
11.015 Million cell updates/sec

Title: US-09-446-109A-5
Perfect score: 100
Sequence: 1 RAARISLGRXYSFKPMPLXR 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	48.0	1676	1 C5HU	complement C5 prec
2	45	45.0	1195	2 B96746	probable kinesin r
3	42	42.0	118	2 S74769	hypothetical prote
4	42	42.0	495	2 T48372	ATB'alpha regulato
5	42	42.0	679	2 T24365	hypothetical prote
6	42	42.0	717	2 T27066	hypothetical prote
7	42	42.0	717	2 T27067	hypothetical prote
8	41	41.0	690	2 T23775	hypothetical prote
9	41	41.0	881	1 A64040	hypothetical prote
10	41	41.0	921	2 T51136	ionotropic glutama
11	41	41.0	923	2 F84732	probable ligand-ga
12	40	40.0	381	2 C70711	hypothetical prote
13	40	40.0	429	2 E64617	hypothetical prote
14	40	40.0	733	2 E82525	primosomal protein
15	39	39.0	74	2 A01268	complement C5 - pi
16	39	39.0	308	2 C75073	hypothetical prote
17	39	39.0	457	2 S03321	regulatory protein
18	39	39.0	508	1 PWZHAM	H+-transporting AT
19	39	39.0	509	1 PWTAM	H+-transporting AT
20	39	39.0	509	1 PWRZAM	H+-transporting AT
21	39	39.0	821	1 D65092	outer membrane ush
22	39	39.0	1011	2 C84524	probable disease r
23	38	38.0	74	2 A25408	complement C5 - bo
24	38	38.0	208	2 B70442	ribosomal protein
25	38	38.0	292	2 D86223	hypothetical prote
26	38	38.0	388	1 B41339	xylose isomerase (
27	38	38.0	388	2 JC1031	xylose isomerase (
28	38	38.0	389	1 TSSMKV	xylose isomerase (
29	38	38.0	394	1 TSSMXR	xylose isomerase (

30	38	38.0	432	2 A28913	regulatory protein
31	38	38.0	515	1 S45583	pisatin demethylas
32	38	38.0	622	2 E69609	cytochrome-c oxida
33	38	38.0	676	1 W2BE43	gene 43 protein -
34	38	38.0	688	2 H96681	protein F1E22.10 (
35	38	38.0	719	2 T52510	hypothetical prote
36	38	38.0	748	2 S59327	hypothetical prote
37	38	38.0	749	2 B86606	primosomal protein
38	38	38.0	749	2 C72018	primosomal protein
39	38	38.0	1426	2 T30817	homeotic protein C
40	37.5	37.5	134	2 D72775	hypothetical prote
41	37	37.0	140	2 A72667	hypothetical prote
42	37	37.0	153	2 S33363	gly96 protein - mo
43	37	37.0	154	1 A33893	superoxide dismuta
44	37	37.0	156	2 JC5537	differentiation-de
45	37	37.0	192	2 A84727	hypothetical prote

ALIGNMENTS

RESULT 1
C5HU
Complement C5 precursor [validated] - human
N:Contains: C5a anaphylatoxin; C5b
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Dec-2000
C:Accession: A40075; A27689; A01267; A01266; S15121
R:Haviland, D.L.; Haviland, J.C.; Fleischer, D.T.; Hunt, R.A.
J. Immunol. 146, 362-368, 1991
A:Title: Complete cDNA sequence of human complement pro-C5. Evidence of truncated tra
A:Reference number: A40075; MUID:91079575
A:Accession: A40075
A:Molecule type: mRNA
A:Residues: 1-1676 <H>
A:Cross-references: GB:M57729; NID:g179982; PIDN:AAA51925.1; PID:g179983
A:Note: 518-Ser was also found
R:Wetzel, R.A.; Lemons, R.S.; Le Beau, M.M.; Barnum, S.R.; Noack, D.; Tack, B.F.
Biochemistry 27, 1474-1482, 1988
A:Title: Molecular analysis of human complement component C5: localization of the str
A:Reference number: A27689; MUID:88209511
A:Accession: A27689
A:Molecule type: mRNA
A:Residues: 412-1676 <WFT>
A:Cross-references: GB:M65134; GB:M18879; NID:g179691; PIDN:AAA51856.1; PID:g179692
R:Fernandez, H.N.; Hugli, T.E.
J. Biol. Chem. 253, 6955-6964, 1978
A:Title: Primary structural analysis of the polypeptide portion of human C5a anaphyla
A:Reference number: A01267; MUID:79005687
A:Accession: A01267
A:Molecule type: protein
A:Residues: 678-751 <FER>
R:Lundvall, A.B.; Wetzel, R.A.; Kristensen, T.; Whitehead, A.S.; Woods, D.E.; Ogden,
J. Biol. Chem. 260, 2108-2112, 1985
A:Title: Isolation and sequence analysis of a cDNA clone encoding the fifth componen
A:Reference number: A01266; MUID:85130937
A:Accession: A01266
A:Molecule type: mRNA
A:Residues: 412-854, 'SLALSPRLCNGKISGHKRLRPGSSDPASASQVAGITGTHHAQPT' <LUN>
A:Cross-references: GB:X02874
A:Note: the carboxyl-terminal part of the sequence in this report appears to be deriv
R:Bohnsack, J.F.; Mollison, K.W.; Buko, A.M.; Ashworth, J.C.; Hill, H.R.
Biochem. J. 273, 635-640, 1991
A:Title: Group B streptococci inactivate complement component C5a by enzymic cleavage
A:Reference number: S15121; MUID:91144547
A:Accession: S15121
A:Contents: annotation
C:Comment: Complement C5 contains two disulfide-linked chains, formed by removal of f
(beta and alpha' chains).
C:Comment: Activation of C5 initiates the spontaneous assembly of the late complement
is the foundation upon which the membrane attack complex is assembled.
C:Comment: C5a has potent spasmogenic and chemotactic activity.
C:Genetics:
A:Gene: GDB:C5

```

A:Cross-references: GDB:119734; OMIM:120900
A:Map position: 9q34.1-9q34.1
C:Superfamily: alpha-2-macroglobulin
C:Keywords: complement alternative pathway; complement pathway; cytolysis; glycoprotein;
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-673/752-1676/Product: complement C5 #status predicted <MAT>
F:19-673/752-1676/Product: C5b #status predicted <C5B>
F:19-673/752-1676/Product: complement C5 and C5b beta chain #status predicted <C5BB>
F:678-1676/Product: complement C5 alpha chain #status predicted <C5A>
F:678-1676/Product: C5a anaphylatoxin #status experimental <C5t>
F:752-1676/Product: C5b alpha chain #status predicted <C5BA>
F:567-810-634-669-698-724-699-731-711-732-866-1527-1101-1159-1375-1505-1405-1474-1520-15
F:741/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:751-752/Cleavage site: Arg-Leu (C5 convertase) #status experimental
F:911-1115-1630/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 48.0%; Score 48; DB 1; Length 1676;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RAARISLGPR 10
Db 714 RAARISLGPR 723

RESULT 2
B96746
Probable kinesin T9N14.6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B96746
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: B96746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1195 <STO>
A:Cross-references: GB:AE005173; NID:g10645376; PIDN:AAG21495.1; GSPDB:GN00141
A:Gene: T9N14.6
A:Map position: 1

Query Match 45.0%; Score 45; DB 2; Length 1195;
Best Local Similarity 58.8%; Pred. No. 13;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 AARISLGPXYSPKMP 18
Db 1073 AKRPPAPRRKSPAPMP 1089

RESULT 3
S74769
hypothetical protein sir1068 - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74769
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis

A:Cross-references: S74322; MUID:97061201
A:Accession: S74769
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-118 <KAN>
A:Cross-references: EMBL:D90901; GB:AB001339; NID:g1651897; PIDN:BAAL16920.1; PID:g165
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synecocystis hypothetical protein sir1068

Query Match 42.0%; Score 42; DB 2; Length 118;
Best Local Similarity 41.2%; Pred. No. 3.5;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 AARISLGPXYSPKMP 18
Db 99 ASRVSYSGKYAFEPNP 115

RESULT 4
T48372
ATB/alpha regulatory subunit of PP2A - Arabidopsis thaliana
N:Alternate names: protein F12E4.240
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48372
R:Revan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lencke,
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24492
A:Accession: T48372
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-495 <BEV>
A:Cross-references: EMBL:AL162751
A:Experimental source: cultivar Columbia; BAC clone F12E4
C:Genetics:
A:Map position: 5
A:Introns: 387/3
A:Note: F12E4.240

Query Match 42.0%; Score 42; DB 2; Length 495;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 AARISLGPXYSPKMP 21
Db 57 AAATTQPPMYSVPEPLPFR 76

RESULT 5
T24365
hypothetical protein T02E1.3b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24365
R:Lennard, N.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19881
A:Accession: T24365
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-679 <WIL>
A:Cross-references: EMBL:281581; PIDN:CAB04666.1; GSPDB:GN00019; CESP:T02E1.3b
A:Experimental source: clone T02E1
C:Genetics:
A:Gene: CESP:T02E1.3b
A:Map position: 1
A:Introns: 17/3; 71/3; 102/3; 214/3; 255/1; 372/3; 398/1; 551/1; 634/1

Query Match 42.0%; Score 42; DB 2; Length 679;

```


OY 3 ARISLGPRXYSFKMP 18
||: || | | | |
Db 297 ARVLGTRSAVFTPLP 312

RESULT 15
A01268
complement C5 - pig (fragment)
N:Contains: C5a anaphylatoxin; C5b
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 29-Jul-1981 #sequence_revision 15-Oct-1994 #text_change 16-Feb-1997
C:Accession: A01268; A26248
R:Gerard, C.; Hugli, T. E.
J. Biol. Chem. 255, 4710-4715, 1980
A:Title: Amino acid sequence of the anaphylatoxin from the fifth component of porcine complement C5
A:Reference number: A01268; MUID:80182137
A:Accession: A01268
A:Molecule type: protein
A:Residues: 1-74 <GER>
R:Gerard, C.; Hugli, T. E.
Proc. Natl. Acad. Sci. U.S.A. 78, 1833-1837, 1981
A:Reference number: A26247; MUID:81199549
A:Contents: annotation; active region
A:Note: although Arg-74 is not essential, residues 72-74 (Leu-Gly-Arg) are required for
R:Zimmermann, B.; Vogt, W.
Hoppe-Seyler's Z. Physiol. Chem. 365, 151-158, 1984
A:Reference number: A26248; MUID:84184201
A:Contents: disulfide bonds
A:Accession: A26248
A:Molecule type: protein
A:Residues: 1-64, 'E', 66-73 <ZIM>
C:Comment: Complement C5 contains two disulfide-linked chains, formed by removal of four (beta and alpha' chains).
C:Comment: Activation of C5 initiates the spontaneous assembly of the late complement cascade. Is the foundation upon which the membrane attack complex is assembled.
C:Comment: C5a has potent spasmogenic and chemotactic activity.
C:Superfamily: alpha-2-macroglobulin
C:Keywords: complement alternate pathway; complement pathway; cytolysis; glycoprotein; 1
F:1-74/Product: C5a anaphylatoxin #status experimental <CST>
F:21-47,22-54,34-55/Disulfide bonds: #status experimental

Query Match 39.08; Score 39; DB 2; Length 74;
Best Local Similarity 70.0%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 RAARISLGPR 10
|||||:
Db 37 RAARIKIGPK 46

Search completed: February 27, 2002, 11:45:16
Job time: 695 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:42 : Search time 78.39 Seconds
(without alignments)
9.822 Million cell updates/sec

Title: US-09-446-109a-5
Perfect score: 100
Sequence: 1 RAARISLGRXYSEKPNPLXR 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	48.0	1676	1 COS_HUMAN	P01031 homo sapien
2	42	42.0	333	1 TRPD_PASMU	P57856 pasteurella
3	41	41.0	881	1 YEBT_HAEIN	P44288 haemophilus
4	40	40.0	381	1 YE88_MYCTU	P71768 mycobacteri
5	39	39.0	74	1 COSA_PIG	P01032 sus scrofa
6	39	39.0	457	1 NTRC_RHOCA	P09432 rhodobacter
7	39	39.0	488	1 ZF92_MOUSE	Q62396 mus musculu
8	39	39.0	508	1 ATP0_MAIZE	P05494 zea mays (m
9	39	39.0	509	1 ATP0_ORYSA	P15998 oryza sativ
10	39	39.0	509	1 ATP0_WHEAT	P12862 triticum ae
11	39	39.0	821	1 Y0IC_ECOLI	P76855 escherichia
12	38	38.0	74	1 COSA_BOVIN	P12082 bos taurus
13	38	38.0	208	1 RS5_AQUAE	Q67563 aquifex aeo
14	38	38.0	386	1 Xyla_STROL	P15587 streptomyc
15	38	38.0	387	1 Xyla_STRRU	P24300 streptomyc
16	38	38.0	388	1 Xyla_STRDI	P50910 streptomyc
17	38	38.0	388	1 Xyla_STRMR	P37031 streptomyc
18	38	38.0	388	1 Xyla_STRVO	P09033 streptomyc
19	38	38.0	390	1 Xyla_STRAL	P24299 streptomyc
20	38	38.0	393	1 Xyla_STRRO	P22857 streptomyc
21	38	38.0	420	1 ARCA_STRCO	Q9rj01 streptomyc
22	38	38.0	432	1 BRLA_EWENT	P10069 emericella
23	38	38.0	515	1 PID9_FUSSO	Q12845 fusarium so
24	38	38.0	622	1 COX1_BACSU	P24010 bacillus su
25	38	38.0	676	1 UL17_VZVD	P09292 varicella-2
26	38	38.0	749	1 PRIA_CHLPN	Q9z6v2 chlamydia p
27	38	38.0	1426	1 CUT2_MOUSE	P70298 mus musculu
28	38	38.0	2124	1 Y192_HUMAN	Q93074 homo sapien
29	37	37.0	153	1 IEX1_MOUSE	P46694 mus musculu
30	37	37.0	154	1 SODC_BRUAB	P15453 bruceella ab
31	37	37.0	156	1 IEX1_HUMAN	P46695 homo sapien
32	37	37.0	217	1 GRP1_RAT	P97576 rattus norv
33	37	37.0	310	1 KHSE_THEAC	Q9hkr6 thermoplasm

34	37	37.0	506	1 ATP0_BETVU	Q06735 beta vulgar
35	37	37.0	507	1 ATP0_BRANA	P22201 brassica na
36	37	37.0	507	1 ATP0_RAPSA	P23413 raphanus sa
37	37	37.0	509	1 ATP0_NICPL	P03495 nicotiana p
38	37	37.0	510	1 ATP0_HELAN	P18260 helianthus
39	37	37.0	707	1 GCVK_HCMVA	P16788 human cytom
40	37	37.0	1044	1 YDE1_SCHPO	Q10435 schizosacch
41	37	37.0	1222	1 PMS_P_HUMAN	Q15155 homo sapien
42	37	37.0	1609	1 YL54_CAEEL	P34434 caenorhabdi
43	37	37.0	1650	1 CA2B_MOUSE	Q64739 mus musculu
44	37	37.0	1736	1 CA2B_HUMAN	P13942 homo sapien
45	36.5	36.5	126	1 YOR3_ADEG1	P20745 avian adeno

ALIGNMENTS

RESULT 1

COS_HUMAN STANDARD; PRT: 1676 AA.

AC P01031;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE COMPLEMENT C5 PRECURSOR [CONTAINS: C5A ANAPHYLATOXIN].

GN C5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=91079575; PubMed=1984448;

RA Haviland D.L., Haviland J.C., Fleischer D.T., Hunt A., Wetsel R.A.;

RT "Complete cDNA sequence of human complement pro-C5. Evidence of

RT truncated transcripts derived from a single copy gene.";

RL J. Immunol. 146:362-368(1991).

RN [2]

RP SEQUENCE OF 412-1676 FROM N.A.

RX MEDLINE=88209511; PubMed=3365401;

RA Wetsel R.A., Lemons R.S., Lebeau M.M., Barnum S.R., Noack D.,

RA Tack B.F.;

RT "Molecular analysis of human complement component C5: localization of

RT the structural gene to chromosome 9.";

RL Biochemistry 27:1474-1482(1988).

RN [3]

RP SEQUENCE OF 412-902 FROM N.A.

RX MEDLINE=85130937; PubMed=2579066;

RA Lundwall A.B., Wetsel R.A., Kristensen T., Whitehead A.S.,

RA Woods D.E., Ogden R.C., Colten H.R., Tack B.F.;

RT "Isolation and sequence analysis of a cDNA clone encoding the fifth

RT complement component.";

RL J. Biol. Chem. 260:2108-2112(1985).

RN [4]

RP SEQUENCE OF 678-751.

RX MEDLINE=79005687; PubMed=690134;

RA Fernandez H.N., Hugli T.E.;

RT "Primary structural analysis of the polypeptide portion of human C5a

RT anaphylatoxin. Polypeptide sequence determination and assignment of

RT the oligosaccharide attachment site in C5a.";

RL J. Biol. Chem. 253:6955-6964(1978).

RN [5]

RP SEQUENCE OF 678-751 FROM N.A.

RX MEDLINE=91144547; PubMed=1996961;

RA Bohnsack J.F., Mollison K.W., Buko A.M., Ashworth J.C., Hill H.R.;

RT "Group B streptococci inactivate complement component C5a by enzymic

RT cleavage at the C-terminus.";

RL Biochem. J. 273:635-640(1991).

RN [6]

RP STRUCTURE BY NMR OF C5A.

RX MEDLINE=88309754; PubMed=3408713;

RA Zwieterweg E.R.P., Mollison K.W., Henkin J., Carter G.W.;

RT "Sequence-specific assignments in the 1H NMR spectrum of the human

RT Inflammatory protein C5a.":
 RL Biochemistry 27:3568-3580(1988).
 RN [7]
 RP STRUCTURE BY NMR OF C5A.
 RA MEDLINE=89207527; PubMed=2784981;
 RX Zaidarweg E.R.P., Nettesheim D.G., Mollison K.W., Carter G.W.:
 "Tertiary structure of human complement component C5a in solution
 from nuclear magnetic resonance data.";
 RL Biochemistry 28:172-185(1989).
 RN [8]
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE=89274164; PubMed=2730871;
 RA Zaidarweg E.R.P., Fesik S.W.:
 "Heteronuclear three-dimensional NMR spectroscopy of the inflammatory
 protein C5a.";
 RL Biochemistry 28:2387-2391(1989).
 RN [9]
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE=97160477; PubMed=9007977;
 RA Zhang X., Boyar W., Galakatos N., Connella N.C.:
 "Solution structure of a unique C5a semi-synthetic antagonist:
 Implications in receptor binding.";
 RL Protein Sci. 6:65-72(1997).
 RN [10]
 RP STRUCTURE BY NMR OF C5A.
 RX MEDLINE=97332508; PubMed=9188742;
 RA Zhang X., Boyar W., Toth M.J., Wengnigle L., Connella N.C.:
 "Structural definition of the C5a C terminus by two-dimensional
 nuclear magnetic resonance spectroscopy.";
 RL Proteins 28:261-267(1997).
 CC -!- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
 CC SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
 CC INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
 CC FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYtic
 CC COMPLEX IS ASSEMBLED.
 CC -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
 CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
 CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
 CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
 CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
 CC POLYMONOPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
 CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
 CC -!- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC
 CC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
 CC BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,
 CC RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA
 CC CHAIN).
 CC -!- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
 CC -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
 CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 855
 CC ONWARD DUE TO THE PRESENCE OF AN ALU REPEAT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M57729; AAA51925.1; "
 CC EMBL: M65134; AAA51856.1; "
 CC PIR: A40075; C5HU.
 CC PIR: S15121; S15121.
 CC PDB: 1KJS; 15-MAY-97.
 CC PDB: 1CFA; 17-SEP-97.
 CC MIM: 120900; "
 CC InterPro: IPR002890; A2M_N.
 CC InterPro: IPR001599; Alpha_2_macroglbln.
 CC InterPro: IPR000020; Anaphylatoxin.
 CC InterPro: IPR001840; Anaphylatoxn.
 CC InterPro: IPR001134; Netrin_C.
 CC Pfam: PF00207; A2M; 1.

DR Pfam: PF01835; A2M_N; 1.
 DR Pfam: PF01821; ANATO; 1.
 DR Pfam: PF01759; NTR; 1.
 DR PRINTS; PRO0004; ANAPHYLATOXN.
 DR SMART; SM00104; ANATO; 1.
 DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; FALSE_NEG.
 DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
 DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
 KW Complement pathway; Complement alternate pathway; Glycoprotein;
 KW Plasma; Membrane attack complex; Cytolysis; Inflammatory response;
 KW Signal; Polymorphism; 3D-structure.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 673 COMPLEMENT C5 BETA CHAIN.
 FT PROPEP 674 677
 FT CHAIN 678 1676 COMPLEMENT C5 ALPHA CHAIN.
 FT PEPTIDE 678 751 C5A ANAPHYLATOXIN.
 FT CHAIN 752 1676 C5B (ALPHA').
 FT DOMAIN 698 732 ANAPHYLATOXIN-LIKE.
 FT DISULFID 698 724
 FT DISULFID 699 731
 FT DISULFID 711 732
 FT CARBOHYD 741 741
 FT CARBOHYD 911 911 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1115 1115 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1630 1630 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 518 518 F -> S.
 FT SEQUENCE 1676 AA; 188331 MW; 87DCAA65FF977D19 CRC64;
 SQ
 Query Match 48.0%; Score 48; DB 1; Length 1676;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RAARISLQPR 10
 DB 714 RAARISLQPR 723
 RESULT 2
 ID TRPD_PASMU STANDARD; PRT; 333 AA.
 AC P57856;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.18).
 GN TRPD OR PM0581.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OC NCBI_TaxID=747;
 OX [1]
 RN [1]
 XP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Faustian M.L., Whittam T.S., Kapur V.:
 "Complete genomic sequence of Pasteurella multocida Pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -!- CATALYTIC ACTIVITY: ANTHRANILATE + PHOSPHORIBOSYLPHOSPHATE =
 CC N-5'-PHOSPHORIBOSYL-ANTHRANILATE + PYROPHOSPHATE.
 CC -!- PATHWAY: SECOND STEP IN BIOSYNTHESIS OF TRYPTOPHAN.
 CC -!- SIMILARITY: BELONGS TO THE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```
DR EMBL: AF006093; AAK02665.1; -
DR InterPro: IPR003262; Anthr_phosphorbsyltransf.
DR InterPro: IPR000312; Glycos_transf_3.
DR InterPro: IPR000053; Thymid_phosphils.
DR Pfam: PF00591; Glycos_transf_3; 1.
DR ProDom: PD001864; Glycos_transf_3; 1.
DR ProDom: PD005916; Thymid_phosphils; 1.
DR Tryptophan biosynthesis; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 333 AA; 35880 MW; F42760CD105F6C5A CRC64;

Query Match: 42.0%; Score 42; DB 1; Length 333;
Best Local Similarity 43.8%; Pred. NO. 4.1; Mismatches 5; Indels 0; Gaps 0;
Matches 7; Conservative 4;

QY 4 R1SLGPRXYSKFPMPL 19
DB 245 RYTLTPQDFGFOPL 260

RESULT 3
YE88_HAEIN STANDARD; PRT; 881 AA.
AC P44288;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN H11672.
GN H11672.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Kerschlmann R.D., Adams W.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerschlmann A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae."
RL Electrophoresis 21:411-429(2000).
CC -!- SIMILARITY: BELONGS TO THE PQ1B FAMILY. STRONG. TO E.COLI YE8T.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32840; AAC23317.1; -
DR TIGR: H11672; -
KW Transmembrane; Complete proteome.
FT TRANSMEM 30 49 POTENTIAL.
SQ SEQUENCE 881 AA; 96677 MW; C278E5822E6A4F55 CRC64;

Query Match: 41.0%; Score 41; DB 1; Length 881;
Best Local Similarity 40.0%; Pred. NO. 17;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 AARISLGPRXYSKFPMPLXR 21
DB 588 AAKIDVSPKGISIQATPLAR 607

RESULT 4
YE88_MYCTU STANDARD; PRT; 381 AA.
AC P71768;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 41-2 KDA PROTEIN RV1488.
GN RV1488 OR MT1533.2 OR MTCY277.09.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva W.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BAND 7 / MEC-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z79701; CAB02038.1; -
DR EMBL: AE007022; AAK45800.1; -
DR TIGR: MT1533.2; -
DR TubercuList; RV1488;
DR InterPro: IPR001107; Band_7.
DR InterPro: IPR001972; Stomatin.
DR Pfam: PF01145; Band_7; 1.
DR SMART: SM00244; PHB; 1.
DR PROSITE: PS01270; BAND_7; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
SQ SEQUENCE 381 AA; 41282 MW; 32747A51D653CF91 CRC64;
```

```
Query Match          40.0%; Score 40; DB 1; Length 381;
Best Local Similarity 47.1%; Pred. No. 11;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 3 ARISLGRXYSEKPMPL 19
    || | | | | | |
DB 63 ARVDLRRVVSPFPQV 79

RESULT 5
COSA_PIG          STANDARD; PRT; 74 AA.
AC P01032;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE COMPLEMENT C5A ANAPHYLATOXIN.
GN C5.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RX MEDLINE=80182137; PubMed=7372604;
RA Gerard C., Hugli T.E.;
RT Amino acid sequence of the anaphylatoxin from the fifth component of
RT porcine complement.;
RL J. Biol. Chem. 255:4710-4715(1980).
RN [2]
RX ACTIVE REGION.
RX MEDLINE=81199549; PubMed=6940191;
RA Gerard C., Hugli T.E.;
RT Identification of classical anaphylatoxin as the des-Arg form of the
RT C5a molecule: evidence of a modulator role for the oligosaccharide
RT unit in human des-Arg74-C5a.;
RL Proc. Natl. Acad. Sci. U.S.A. 78:1833-1837(1981).
RN [3]
RX STRUCTURE BY NMR.
RX MEDLINE=90248365; PubMed=2337573;
RA Williamson M.P., Madison V.S.;
RT "Three-dimensional structure of porcine C5adesArg from 1H nuclear
RT magnetic resonance data.";
RL Biochemistry 29:2895-2905(1990).
CC -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC POLYMPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
DR PIR; A01268; C5PGAT.
DR PDB; 1CSA; 15-OCT-91.
DR InterPro; IPR001599; Alpha_2_macroglobulin.
DR InterPro; IPR000020; Anaphylatoxin.
DR InterPro; IPR001840; Anaphylatoxn.
DR Pfam; PF01821; ANATO; 1.
DR PRINTS; PR00004; ANAPHYLATOXN.
DR SMART; SM00104; ANATO; 1.
DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; PARTIAL.
DR PROSITE; PS01177; ANAPHYLATOXIN_1; 1.
DR PROSITE; PS01178; ANAPHYLATOXIN_2; 1.
KW Complement pathway; Complement alternate pathway; Plasma;
KW Inflammatory response; 3d-structure.
FT DOMAIN 21 55 ANAPHYLATOXIN-LIKE.
FT DISULFID 21 47
FT DISULFID 22 54
FT DISULFID 34 55
FT SITE 72 74
FT FT (ARE REQUIRED FOR 90% OF C5A ACTIVITY
FT FT (ALTHOUGH ARG-74 IS NOT ESSENTIAL).
FT HELIX 2 11
FT TURN 13 14

FT HELIX 16 26
FT HELIX 34 40
FT HELIX 45 62
FT TURN 63 64
SQ SEQUENCE 74 AA; 8609 MW; 11AAF2E94A026EB3 CRC64;

Query Match          39.0%; Score 39; DB 1; Length 74;
Best Local Similarity 70.0%; Pred. No. 2.9;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RAARISLGRP 10
    ||||| :||:
DB 37 RAARIKIGPK 46

RESULT 6
NTRC_RHOCA
ID NTRC_RHOCA STANDARD; PRT; 457 AA.
AC P09432;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE NITROGEN REGULATION PROTEIN NTRC.
GN NTRC OR NIFRI.
OS Rhodobacter capsulatus (Rhodospseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN=SB1003.
RX MEDLINE=89218961; PubMed=2710108;
RA Jones R., Haselkorn R.;
RT "The DNA sequence of the Rhodobacter capsulatus ntrA, ntrB and ntrC
RT gene analogues required for nitrogen fixation.";
RL Mol. Gen. Genet. 215:507-516(1989).
RN [2]
RX SEQUENCE FROM N.A.
RX STRAIN=SB1003;
RX MEDLINE=93360820; PubMed=8355615;
RA Foster-Hartnett D., Cullen P.J., Gabbett K.K., Kranz R.G.;
RT "Sequence, genetic, and lacZ fusion analyses of a nifR3-ntrB-ntrC
RT operon in Rhodobacter capsulatus.";
RL Mol. Microbiol. 8:903-914(1993).
CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM NTRB/NTRC
CC INVOLVED IN THE ACTIVATION OF NITROGEN ASSIMILATORY GENES SUCH AS
CC GLNA. NTRC IS PHOSPHORYLATED BY NTRB AND INTERACTS WITH SIGMA-54.
CC -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
CC -!- SIMILARITY: THE CENTRAL REGION CONTAINS A SIGMA-54 FACTOR
CC INTERACTION ATP-BINDING DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X12359; CAA30922.1; -
DR EMBL; X72382; CAA51075.1; -
DR PIR; S03321; S03321.
DR HSSP; P11028; 3FIS.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR002197; HTH_Fis.
DR InterPro; IPR001789; Response_reg.
DR InterPro; IPR002078; Sig54_interact.
DR Pfam; PF00072; response_reg; 1.
DR Pfam; PF00158; sigma54; 1.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00448; REC; 1.
```

```

DR PROSITE: PS00675: SIGMA54_INTERACT_1; 1.
DR PROSITE: PS00676: SIGMA54_INTERACT_2; FALSE_NEG.
DR PROSITE: PS00688: SIGMA54_INTERACT_3; 1.
DR PROSITE: PS00689: SIGMA54_INTERACT_4; 1.
KW Nitrogen fixation: Transcription regulation: Repressor: Activator;
KW DNA-binding: ATP-binding: Phosphorylation: Sensor: transduction.
FT MOD_RES 53 53 PHOSPHORYLATION (BY SIMILARITY).
FT DOMAIN 143 343 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
FT NP_BIND 171 178 ATP (POTENTIAL).
FT DNA_BIND 425 444 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 457 AA; 50042 MW; 7EB04B4FE4740306 CRC64;

Query Match: 39.08; Score 39; DB 1; Length 457;
Best Local Similarity 47.6%; Pred. No. 20;
Matches 10; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 RAARISLGRPRXYSEKPMPLXR 21
   |||||:|:|:|:|
DB 303 RAERDGLGMRFAFSAEAMGLVR 323

RESULT 7
ID 2F92_MOUSE STANDARD; PRT; 488 AA.
AC Q62396;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ZINC FINGER PROTEIN 92 (ZFP-92).
GN ZFP92.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=96425694; PubMed=8828036;
RA Levin M.L., Chatterjee A., Pragliola A., Worley K.C., Wehnert M.,
RA Zhuchenko O., Smith R.F., Lee C.C., Herman G.E.;
RT "A comparative transcription map of the murine bare patches (Bpa) and
RT striated (Str) critical regions and human Xq28.";
RL Genome Res. 6:465-477(1996).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U47104; AAC52629.1;
CC MGD: MGI:108094; Zfp92.
CC InterPro: IPR001909; KRAB.
CC Pfam: PF01352; KRAB; 1.
CC Pfam: PF00096; zf-C2H2; 9.
CC PRINTS: PR00048; ZINCFINGER.
CC SMART: SM00349; KRAB; 1.
CC SMART: SM00355; Znf_C2H2; 9.
CC PROSITE: PS00805; KRAB; 1.
CC PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.
CC PROSITE: PS0157; ZINC_FINGER_C2H2_2; 9.
KW Transcription regulation: DNA-binding; Zinc-finger; Metal-binding;
KW Nuclear protein; Repeat.
FT DOMAIN 14 85 KRAB.

```

```

FT DOMAIN 141 432 ZINC FINGERS.
FT ZN_FING 141 163 C2H2-TYPE.
FT ZN_FING 169 191 C2H2-TYPE.
FT ZN_FING 197 219 C2H2-TYPE.
FT ZN_FING 225 247 C2H2-TYPE.
FT ZN_FING 253 275 C2H2-TYPE.
FT ZN_FING 281 303 C2H2-TYPE.
FT ZN_FING 337 359 C2H2-TYPE.
FT ZN_FING 410 432 C2H2-TYPE.
SQ SEQUENCE 488 AA; 55961 MW; 6A649E30F2043699 CRC64;

Query Match: 39.08; Score 39; DB 1; Length 488;
Best Local Similarity 42.1%; Pred. No. 21;
Matches 8; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 3 ARISLGRPRXYSEKPMPLXR 21
   |||||:|:|:|:|
DB 112 SRSLQRPHDFRPNPIVR 130

RESULT 8
ID ATP0_MAIZE STANDARD; PRT; 508 AA.
AC P05494;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL (EC 3.6.1.34).
GN ATPA.
OS Zea mays (Maize).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxID=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Braun C.J., Levings C.S. III;
RT "Nucleotide sequence of the Fl-ATPase alpha subunit from maize
RT mitochondria.";
RL Plant Physiol. 79:571-577(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88311189; PubMed=2900697;
RA Isaac P.G., Brennick A., Dunbar S.M., Leaver C.J.;
RT "The mitochondrial genome of fertile maize (Zea mays L.) contains two
RT copies of the gene encoding the alpha-subunit of the Fl-ATPase.";
RL Curr. Genet. 10:321-328(1985).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE ALPHA CHAIN IS A REGULATORY
CC SUBUNIT.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M16222; AAA70269.1;
CC EMBL: Z00026; CAA77319.1;
CC PIR: A23757; PFMZAM.
CC HSSP: P19483; ICOW.
CC MaizDB: 69198;
CC InterPro: IPR000790; ATPase_A_C.

```

```

DR InterPro: IPR000194; ATPase_alpha_beta.
DR Pfam: PF00422; ATP-synt_A-C; 1.
DR Pfam: PF00006; ATP-synt_ab; 1.
DR ProDom: PD001099; ATPase_A-C; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW ATP synthesis: CF(1); Hydrogen ion transport;
KW Hydrolyase: ATP-binding; Mitochondrion.
FT NP_BIND 171 178 ATP (BY SIMILARITY).
FT ACT_SITE 373 373
SQ SEQUENCE 508 AA; 55180 MW; FBA3DD33141A0456 CRC64;

Query Match 39.0%; Score 39; DB 1; Length 508;
Best Local Similarity 28.6%; Pred. No. 22;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 RAARISLGPRXYSEKPMPLXR 21
| | | | | : : : : :
Db 423 RGA RTEVPKPOQEPLPIEK 443

RESULT 9
ATPO_ORYSA STANDARD; PRT; 509 AA.
AC P15998;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL (EC 3.6.1.34).
GN ATPA.
OS Oryza sativa (Rice).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, JAPONICA; TISSUE=Shoot;
RX MEDLINE=90206808; PubMed=2138730;
RA Kadowaki K.I., Boireau P., Laporte J.;
RT "Nucleotide sequence of the F1-ATPase alpha subunit gene from rice
RT Mitochondria.";
RL Nucleic Acids Res. 18:1302-1302(1990).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE ALPHA CHAIN IS A REGULATORY
CC SUBUNIT.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X51422; CAA35787.1;
CC PIR: JQ0411; PWRZAM.
CC HSSP: P19483; ICOW.
CC InterPro: IPR000790; ATPase_A-C.
CC InterPro: IPR000194; ATPase_alpha_beta.
CC Pfam: PF00422; ATP-synt_A-C; 1.
CC Pfam: PF00006; ATP-synt_ab; 1.
CC ProDom: PD001099; ATPase_A-C; 1.
CC PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW ATP synthesis: CF(1); Hydrogen ion transport;
KW Hydrolyase: ATP-binding; Mitochondrion.
FT NP_BIND 171 178 ATP (BY SIMILARITY).
FT ACT_SITE 373 373
SQ SEQUENCE 509 AA; 55264 MW; 2BD7893B255EF66B CRC64;

Query Match 39.0%; Score 39; DB 1; Length 509;
Best Local Similarity 28.6%; Pred. No. 22;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

FT NP_BIND 171 178 ATP (BY SIMILARITY).
FT ACT_SITE 373 373
SQ SEQUENCE 509 AA; 55281 MW; 7E6C7561B77C0668 CRC64;

Query Match 39.0%; Score 39; DB 1; Length 509;
Best Local Similarity 28.6%; Pred. No. 22;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 RAARISLGPRXYSEKPMPLXR 21
| | | | | : : : : :
Db 423 RGA RTEVPKPOQEPLPIEK 443

RESULT 10
ATPO_WHEAT STANDARD; PRT; 509 AA.
AC P12862;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL (EC 3.6.1.34).
GN ATPA.
OS Triticum aestivum (Wheat).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Triticeae; Triticum.
OX NCBI_TaxID=4565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90016824; PubMed=2529479;
RA Schulte E., Staubach S., Laser B., Kueck U.;
RT "Wheat mitochondrial DNA: organization and sequences of the atpA and
RT atp9 genes.";
RL Nucleic Acids Res. 17:7531-7531(1989).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE ALPHA CHAIN IS A REGULATORY
CC SUBUNIT.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE ATPASE ALPHA/BETA CHAINS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X15918; CAA34060.1;
CC PIR: S06007; PWMZAM.
CC HSSP: P19483; ICOW.
CC InterPro: IPR000790; ATPase_A-C.
CC InterPro: IPR000194; ATPase_alpha_beta.
CC Pfam: PF00422; ATP-synt_A-C; 1.
CC Pfam: PF00006; ATP-synt_ab; 1.
CC ProDom: PD001099; ATPase_A-C; 1.
CC PROSITE: PS00152; ATPASE_ALPHA_BETA; 1.
KW ATP synthesis: CF(1); Hydrogen ion transport;
KW Hydrolyase: ATP-binding; Mitochondrion.
FT NP_BIND 171 178 ATP (BY SIMILARITY).
FT ACT_SITE 373 373
SQ SEQUENCE 509 AA; 55264 MW; 2BD7893B255EF66B CRC64;

Query Match 39.0%; Score 39; DB 1; Length 509;
Best Local Similarity 28.6%; Pred. No. 22;
Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

FT NP_BIND 171 178 ATP (BY SIMILARITY).
FT ACT_SITE 373 373
SQ SEQUENCE 509 AA; 55281 MW; 7E6C7561B77C0668 CRC64;

```

```

QY 1 RAARISLGPXYSEKPMPLXR 21
Db 423 RGAALTEVPKQOYEPLPIEK 443

RESULT 11
YQIG_ECOLI STANDARD; PRT; 821 AA.
ID YQIG_ECOLI
AC P76655;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN YQIG PRECURSOR.
GN YQIG OR B3046.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Coeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF A FIMBRIAL
CC SUBUNIT ACROSS THE OUTER MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000386; AAC76082.1; .
DR EcoGene: EGI4228; yqig.
DR InterPro: IPR000015; Fimb_usher.
DR Pfam: PF00577; Usher; 1.
DR PROSITE: PS01151; FIMBRIAL_USHER; 1.
KW Hypothetical protein; Outer membrane; Transmembrane; Fimbria;
KW Transport; Signal; Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 821 HYPOTHETICAL OUTER MEMBRANE USHER PROTEIN
FT YQIG.
FT DISULFID 798 817 POTENTIAL.
FT SEQUENCE 821 AA; 91886 MW; C4C9BCC9DA210286 CRC64;

Query Match 39.0%; Score 39; DB 1; Length 821;
Best Local Similarity 60.0%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 9 PRXYSEKPM 18
Db 217 PRYLFRPIP 226

RESULT 12
COSA_BOVIN STANDARD; PRT; 74 AA.
ID COSA_BOVIN
AC P12082;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)

COMPLEMENT C5A ANAPHYLATOXIN.
C5.
Bos taurus (Bovine).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=86136134; PubMed=3081348;
RA Gennaro R., Simonic T., Negri A., Mottola C., Secchi C., Ronchi S.,
RA Romeo D.;
RT "C5a fragment of bovine complement. Purification, bioassays,
RT amino-acid sequence and other structural studies.";
RL Eur. J. Biochem. 155:77-86(1986).
RN [2]
RP STRUCTURE BY NMR.
RX MEDLINE=89005703; PubMed=3262536;
RA Zarbock J., Gennaro R., Romeo D., Clore G.M., Gronenborn A.M.;
RT "A proton nuclear magnetic resonance study of the conformation of
RT bovine anaphylatoxin C5a in solution.";
RL FEBS Lett. 238:289-294(1988).
CC -1- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC -1- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
DR PIR: A25408; A25408.
DR HSSP: P01032; IC5A.
DR InterPro: IPR001599; Alpha_2_macrogloblin.
DR InterPro: IPR000020; Anaphylatoxin.
DR InterPro: IPR001840; Anaphylatoxin.
DR Pfam: PF01821; ANATO; 1.
DR PRINTS: PR00004; ANAPHYLATOXN.
DR SMART: SM00104; ANATO; 1.
DR PROSITE: PS00477; ALPHA_2_MACROGLOBULIN; PARTIAL.
DR PROSITE: PS01177; ANAPHYLATOXIN_1; 1.
DR PROSITE: PS01178; ANAPHYLATOXIN_2; 1.
KW Complement pathway; Complement alternate pathway; Plasma;
KW Inflammatory response.
FT DOMAIN 21 55 ANAPHYLATOXIN-LIKE.
FT DISULFID 21 47 BY SIMILARITY.
FT DISULFID 22 54 BY SIMILARITY.
FT DISULFID 34 55 BY SIMILARITY.
FT SEQUENCE 74 AA; 8517 MW; C09DF742D12D70F6 CRC64;

Query Match 38.0%; Score 38; DB 1; Length 74;
Best Local Similarity 77.8%; Pred. No. 4.3;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RAARISLGP 9
Db 37 RAARIAIGP 45

RESULT 13
RSS_AQUAE STANDARD; PRT; 208 AA.
ID RSS_AQUAE
AC O67563;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S5.
GN RPE OR AQ_1645.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.

```


QY 8 GPRXYSKP 16
!!! : !!!
Db 281 GPRHDFKP 289

RESULT 15
XYLA_STRRU STANDARD; PRT: 387 AA.
ID XYLA_STRRU
AC P24300;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE XYLOSE ISOMERASE (EC 5.3.1.5).
GN XYLA.
OS Streptomyces rubiginosus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1929;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92041569; PubMed=1657868;
RA Wong H.C., Ting Y., Lin H.C., Reichert F., Myambo K., Watt K.W.,
RA Toy P.L., Drummond R.J.;
RT "Genetic organization and regulation of the xylose degradation genes
in Streptomyces rubiginosus";
RL J. Bacteriol. 173:6849-6858(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=9282788; PubMed=2734296;
RA Carrell H.L., Glusker J.P., Burger V., Manfre F., Tritsch D.,
RA Biellmann J.-F.;
RT "X-ray analysis of D-xylose isomerase at 1.9 A: native enzyme in
complex with substrate and with a mechanism-designed inactivator";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4440-4444(1989).
CC -!- FUNCTION: INVOLVED IN D-XYLOSE CATABOLISM.
CC -!- CATALYTIC ACTIVITY: D-XYLOSE -> D-XYLULOSE.
CC -!- COFACTOR: MAGNESIUM IS NECESSARY FOR THE ACTIVITY.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE XYLOSE ISOMERASE FAMILY.
CC -!- CAUTION: ACCORDING TO THE CRYSTALLOGRAPHIC STUDY RESIDUE 40
COULD BE GLN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M73789; AAA26838.1;
DR PIR; B41339;
DR PDB; 1XIS: 15-JUL-92.
DR PDB; 2XIS: 15-JUL-92.
DR PDB; 3XIS: 15-JUL-92.
DR PDB; 4XIS: 15-JUL-92.
DR PDB; 1XIB: 22-JUN-94.
DR PDB; 1XIC: 22-JUN-94.
DR PDB; 1XID: 22-JUN-94.
DR PDB; 1XIE: 22-JUN-94.
DR PDB; 1XIF: 22-JUN-94.
DR PDB; 1XIG: 22-JUN-94.
DR PDB; 1XIH: 22-JUN-94.
DR PDB; 1XII: 22-JUN-94.
DR PDB; 1XIJ: 22-JUN-94.
DR PDB; 8XIA: 15-OCT-91.
DR PDB; 9XIA: 15-JUL-92.
DR InterPro; IPR001998; Xylose_isom.
CR Pfam; PF00259; Xylose_isom; 1.
DR PRINTS; PR00688; XYLOSEISOMRASE.
DR PROSITE; PS00172; XYLOSE_ISOMERASE_1; 1.
DR PROSITE; PS00173; XYLOSE_ISOMERASE_2; 1.

Isomerase; Pentose shunt; Xylose metabolism; Magnesium; 3D-structure.
NW
FT INIT_MET 0
FT ACT_SITE 53
FT METAL 180
FT METAL 216
FT METAL 244
FT METAL 286
FT HELIX 6
FT STRAND 10
FT HELIX 14
FT TURN 18
FT STRAND 23
FT TURN 24
FT STRAND 26
FT HELIX 35
FT TURN 45
FT STRAND 49
FT HELIX 54
FT TURN 60
FT HELIX 64
FT TURN 82
FT STRAND 84
FT STRAND 87
FT HELIX 96
FT TURN 99
FT TURN 103
FT HELIX 108
FT TURN 128
FT STRAND 132
FT TURN 138
FT STRAND 141
FT HELIX 145
FT TURN 172
FT STRAND 176
FT STRAND 189
FT HELIX 195
FT TURN 203
FT HELIX 208
FT STRAND 211
FT STRAND 216
FT HELIX 217
FT TURN 222
FT HELIX 227
FT TURN 237
FT STRAND 240
FT STRAND 244
FT STRAND 247
FT STRAND 257
FT TURN 259
FT HELIX 264
FT TURN 277
FT STRAND 283
FT TURN 291
FT HELIX 295
FT TURN 323
FT HELIX 332
FT TURN 334
FT TURN 338
FT TURN 343
FT HELIX 345
FT TURN 353
FT TURN 356
FT TURN 361
FT TURN 366
FT HELIX 371
FT TURN 384
SQ SEQUENCE 387 AA; 43096 MW; 21CD34271AB85EE3 CRC64;

Query Match 38.08; Score 38; DB 1; Length 387;
Best Local Similarity 66.74; Pred. No. 25;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Caps 0;

Oy 8 GPRXYSPKP 16
III : III
Db 281 GPRHDFKP 289

Search completed: February 27, 2002, 11:42:43
Job time: 542 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:04 ; Search time 281.76 Seconds
(without alignments)
10.902 Million cell updates/sec

Title: US-09-446-109A-5
perfect score: 100
Sequence: 1 RAARISLGPXXYSFKPMPLXR 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_17:*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_unclassified.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	%		Length	DB	ID	Description
			Match	Length				
1		45	45.0	312	11	Q99LP8	Q99LP8 mus musculu	
2		45	45.0	312	11	Q99LF3	Q99LF3 mus musculu	
3		45	45.0	1195	10	Q99C70	Q99C70 arabidopsis	
4		42.5	42.5	462	5	Q9GRM7	Q9GRM7 leishmania	
5		42	42.0	118	2	P72904	P72904 synechocyst	
6		42	42.0	350	3	Q00870	Q00870 fusarium so	
7		42	42.0	495	10	Q04375	Q04375 arabidopsis	
8		42	42.0	717	5	Q45938	Q45938 caenorhabdi	
9		42	42.0	717	5	Q45939	Q45939 caenorhabdi	
10		42	42.0	947	3	Q9P449	Q9P449 schizophyll	
11		42	42.0	1616	4	Q15034	Q15034 homo sapien	
12		41	41.0	411	8	Q9T719	Q9T719 lactoris fe	
13		41	41.0	550	10	Q9LLM4	Q9LLM4 eucalyptus	
14		41	41.0	690	5	Q9XUC6	Q9XUC6 caenorhabdi	
15		41	41.0	921	10	Q9SDQ4	Q9SDQ4 arabidopsis	
16		41	41.0	923	10	Q9ZV68	Q9ZV68 arabidopsis	
17		41	41.0	1473	10	Q9SDH1	Q9SDH1 oryza sativa	
18		40	40.0	266	2	O07405	O07405 mycobacteri	
19		40	40.0	374	2	Q9CBM5	Q9CBM5 mycobacteri	

Query Match	45.08;	Score 45;	DB 11;	Length 312;
Best Local Similarity	47.18;	Pred. No. 7.4;		

QY	3	ARISLGRPRXYSEKPMPL	19
.		: :	: :
Db	268	ALLALGCKLYFFFOIPL	284

RESULT 2
Q99LF3
ID Q99LF3
PRELIMINARY;
PRT: 312 AA.

NC	Q59UR3	01-JUN-2001	(TEMBLrel. 17, Created)
DT		01-JUN-2001	(TEMBLrel. 17, Last sequence update)
DT		01-JUN-2001	(TEMBLrel. 17, Last annotation update)
DE			SIMILAR TO HYPOTHETICAL PROTEIN MGC3262.
OS			Mus musculus (Mouse)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI TaxID=10090;

ALIGNMENTS

RESULT	1
--------	---

Q99LP8
ID Q99LP8 PRELIMINARY; PRT; 312 AA.

AC Q99LP8: 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE DE SIMILAR TO HYPOTHETICAL PROTEIN.

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RP	SEQUENCE FROM N.A.
RC	TISSUE=MAMMARY TUMOR;
RA	Strausberg R.;
kL	Submitted (JAN-2001) to the ENBL/GenBank/DBJ databases.
EMBL	BC002283; AAH02282.1; -
DR	SEQUENCE 312 AA; 34900 MW; 36448819FF529F7D CRC64;

RN [1]
RP SEQUENCE FROM N.A.

RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003289; AA03289.1;
SQ SEQUENCE 312 AA; 34934 MW; 66E8B2B3FF5FF70 CRC64;

Query Match 45.0%; Score 45; DB 11; Length 312;
Best Local Similarity 47.1%; Pred. No. 7.4;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Caps 0;

Qy 3 ARLSLGPRXYSEFKPMP 19
I : : I : : I : : I : : I : :
Db 268 ALLALGCKLYEFOPLPL 284

RESULT 3
Q9C7T0 PRELIMINARY; PRT; 1195 AA.

AC Q9C7T0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KINESIN, PUTATIVE.
GN T9N14.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.D., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egtu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Millitscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
DR EMBL; AC067754; AAC51794.1;
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; P00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00067; KINESIN_MOTOR_DOMAIN2; 1.
KW ATP-binding; Coiled coil; Microtubules; Motor protein.
SQ SEQUENCE 1195 AA; 133383 MW; 3DA8DF7318451F8 CRC64;

Query Match 45.0%; Score 45; DB 10; Length 1195;
Best Local Similarity 58.8%; Pred. No. 32;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Caps 0;

Qy 2 AARISLGRXYSEFKPMP 18
I : : I : : I : : I : : I : :
Db 1073 AKRIPPAPRRKSEAPMP 1089

RESULT 4
Q9GRM7

ID Q9GRM7 PRELIMINARY; PRT; 462 AA.
AC Q9GRM7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE POSSIBLE HYPOTHETICAL 140.4 KD PROTEIN.
GN P214.05.
OS Leishmania major.
OC Eukaryota; Euzoenzoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Saunders D., Murphy L., Harris D., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL449144; CAC14611.1;
SQ SEQUENCE 462 AA; 50128 MW; BF93BF8B907A7DD0 CRC64;

Query Match 42.5%; Score 42.5; DB 5; Length 462;
Best Local Similarity 50.0%; Pred. No. 31;
Matches 11; Conservative 2; Mismatches 8; Indels 1; Caps 1;

Qy 1 RAARISL-GPRXYSEFKPMP 21
I : : I : : I : : I : : I : :
Db 134 RSAEISLYGPANADFPQMQLTK 155

RESULT 5
P72904

ID P72904 PRELIMINARY; PRT; 118 AA.
AC P72904;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 13.7 KDA PROTEIN.
GN SUR1068.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90901; BAA16920.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 118 AA; 13731 MW; D384D5E6826D903C CRC64;

Query Match 42.0%; Score 42; DB 2; Length 118;
Best Local Similarity 41.2%; Pred. No. 8.7;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Caps 0;

Qy 2 AARISLGRXYSEFKPMP 18

DR Mendel; 17339; Arath; 2799; 17339.

CC -1- SIMILARITY: TO ATP SYNTHASE ALPHA CHAIN, C TERMINUS.
 DR EMBL: AF197710; AAF17042.1; -
 DR InterPro: IPR000194; ATPase_alpha_beta.
 DR InterPro: IPR000790; ATPase_A_C.
 DR Pfam: PF004422; ATP-synt_A-c; 1.
 DR Pfam: PF00006; ATP-synt_ab; 1.
 DR ProDom: PD001099; ATPase_A_C; 1.
 DR ProSITE: PS00152; ATPase_ALPHA_BETA; 1.
 KW ATP synthesis: ATP-binding; CF(1); Hydrogen ion transport; Hydrolase;
 KW Mitochondrion.
 FT NON_TER 1
 FT NON_TER 411 411
 SQ SEQUENCE 411 AA; 44302 MW; 043EBA65BF984C1 CRC64;

Query Match 41.0%; Score 41; DB 8; Length 411;
 Best Local Similarity 28.6%; Pred. No. 50;
 Matches 6; Conservative

Matches 6; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 RAARISLGPRXYSFKPMPLXR 21

Db 389 RGA RTEVPKQQQYEPPIEK 409

RESULT 13
 Q9LLM4 PRELIMINARY; PRT; 550 AA.
 ID Q9LLM4
 AC Q9LLM4

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE K+-NA+ SYMPORTER HKT1.

OS Eucalyptus camaldulensis (Murray red gum).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Myrtales; Myrtaceae; Eucalyptus.
 OX NCBI_TaxID=34316;

RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=YOUNG STEM.
 RA Fairbairn D.J., Gomez-Gallego S., Day S., Sawbridge T., Teasdale R.D.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF176035; AAF9728.1; -

DR InterPro: IPR003445; TrkH.

DR Pfam: PF02386; TrkH; 1.

SQ SEQUENCE 550 AA; 61950 MW; 4BADFCAE2BC85441 CRC64;

Query Match 41.0%; Score 41; DB 10; Length 550;
 Best Local Similarity 57.1%; Pred. No. 69;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 6 SLGPRXYSFKPMPL 19

Db 68 ALGPRDTSFPRDL 81

RESULT 14

Q9XUC6 PRELIMINARY; PRT; 690 AA.
 ID Q9XUC6
 AC Q9XUC6

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE M162.7 PROTEIN.

GN M162.7.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Basham V.;

Submittted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 [2]

SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilsson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,

RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;

RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.;

RL Nature 368:32-38(1994).

DR EMBL: 282278; CAB05258.1; -

DR InterPro: IPR003125; WSN.

DR Pfam: PF02206; WSN; 1.

DR SMART: SM00453; WSN; 1.

SQ SEQUENCE 690 AA; 78360 MW; 0E9A2B2A2547F124 CRC64;

Query Match 41.0%; Score 41; DB 5; Length 690;
 Best Local Similarity 47.4%; Pred. No. 88;
 Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 RAARISLGPRXYSFKPMPL 19

Db 311 KVKRASLDNLASSFKPLPL 329

RESULT 15

Q9SDQ4 PRELIMINARY; PRT; 921 AA.

ID Q9SDQ4

AC Q9SDQ4

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GLR5

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. COLUMBIA; TISSUE=WHOLE SEEDLINGS AT LEAF STAGE 2;

RA Davenport R.J., Kiegle E.A., Tester M.;

RT *GLR5, an ionotropic glutamate receptor ortholog from Arabidopsis.;

RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF210701; AAF21042.1; -

DR InterPro: IPR001311; SBP_glut_receptor.

DR InterPro: IPR001320; Ion_glut_receptor.

DR InterPro: IPR001828; ANF_receptor.

DR Pfam: PF00060; lig_chan; 1.

DR Pfam: PF01094; ANF_receptor; 1.

DR SMART: SM00079; DBPe; 1.

SQ SEQUENCE 921 AA; 103491 MW; 69F9707A4D63C55D CRC64;

Query Match 41.0%; Score 41; DB 10; Length 921;
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;
 Matches 10; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

QY 6 SLGPRXY--SFKPMPLXR 21

Db 218 SIGPRVILHFGPDPLR 235

Job time: 985 sec

100

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:02 : Search time 132.19 Seconds
(without alignments)
3.575 Million cell updates/sec

Title: US-09-446-109A-5
Perfect score: 100
Sequence: 1 RAARISIGPRXYSFPMPLXR 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	48.0	74	1 US-08-463-224-3	Sequence 3, Appli
2	48	48.0	74	2 US-08-463-177-3	Sequence 3, Appli
3	48	48.0	74	4 US-09-246-500B-10	Sequence 10, Appli
4	42	42.0	10	1 US-08-299-285-7	Sequence 7, Appli
5	42	42.0	10	1 US-08-299-285-10	Sequence 10, Appli
6	42	42.0	10	1 US-08-299-285-24	Sequence 24, Appli
7	42	42.0	10	2 US-08-985-126-7	Sequence 7, Appli
8	42	42.0	10	2 US-08-985-126-10	Sequence 10, Appli
9	42	42.0	10	2 US-08-985-126-24	Sequence 24, Appli
10	42	42.0	10	5 PCT-US95-11126-7	Sequence 7, Appli
11	42	42.0	10	5 PCT-US95-11126-10	Sequence 10, Appli
12	42	42.0	10	5 PCT-US95-11126-24	Sequence 24, Appli
13	41	41.0	10	1 US-08-299-285-5	Sequence 5, Appli
14	41	41.0	10	1 US-08-299-285-13	Sequence 13, Appli
15	41	41.0	10	2 US-08-985-126-5	Sequence 5, Appli
16	41	41.0	10	2 US-08-985-126-13	Sequence 13, Appli
17	41	41.0	10	5 PCT-US95-11126-5	Sequence 5, Appli
18	41	41.0	10	5 PCT-US95-11126-13	Sequence 13, Appli
19	40	40.0	10	1 US-08-299-285-14	Sequence 14, Appli
20	40	40.0	10	1 US-08-299-285-23	Sequence 23, Appli
21	40	40.0	10	2 US-08-985-126-14	Sequence 14, Appli
22	40	40.0	10	2 US-08-985-126-23	Sequence 23, Appli
23	40	40.0	10	5 PCT-US95-11126-14	Sequence 14, Appli
24	40	40.0	10	5 PCT-US95-11126-23	Sequence 23, Appli
25	40	40.0	77	4 US-09-246-500B-14	Sequence 14, Appli
26	40	40.0	381	1 US-08-781-562-6	Sequence 6, Appli
27	39	39.0	74	4 US-09-246-500B-11	Sequence 11, Appli

28	38	38.0	74	4 US-09-246-500B-12	Sequence..
29	38	38.0	347	1 US-07-637-870-5	Sequence 10,
30	38	38.0	347	1 US-07-640-476-10	Sequence 8, Appli
31	38	38.0	348	1 US-07-637-399-8	Sequence 8, Appli
32	38	38.0	348	1 US-08-112-703-8	Sequence 3, Appli
33	38	38.0	387	1 US-07-637-870-3	Sequence 9, Appli
34	38	38.0	387	1 US-07-637-399-9	Sequence 12, Appli
35	38	38.0	387	1 US-07-640-476-12	Sequence 9, Appli
36	38	38.0	387	1 US-08-112-703-9	Sequence 4, Appli
37	38	38.0	388	1 US-07-637-870-4	Sequence 7, Appli
38	38	38.0	388	1 US-07-637-399-7	Sequence 7, Appli
39	38	38.0	388	1 US-07-640-476-7	Sequence 7, Appli
40	38	38.0	388	1 US-08-112-703-7	Sequence 11, Appli
41	38	38.0	389	1 US-07-640-476-11	Sequence 12, Appli
42	37	37.0	10	1 US-08-299-285-12	Sequence 12, Appli
43	37	37.0	10	2 US-08-985-126-12	Sequence 12, Appli
44	37	37.0	10	5 PCT-US95-11126-12	Sequence 1, Appli
45	37	37.0	154	1 US-07-641-346B-1	

ALIGNMENTS

RESULT 1
US-08-463-224-3
: Sequence 3, Application US/08463224
: Patent No. 5807824
: GENERAL INFORMATION:
: APPLICANT: van Oostrum, Jan
: APPLICANT: Boyar, William C.
: APPLICANT: Galakatos, Nicholas G.
: APPLICANT: Schmitz, Albert
: APPLICANT: van Heeke, Gino
: TITLE OF INVENTION: C5a Receptor Antagonists Having
: SUBSTANTIALLY No. 5807824 Agonist Activity
: NUMBER OF SEQUENCES: 67
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
: STREET: 600 South Avenue West
: CITY: Westfield
: STATE: NJ
: COUNTRY: USA
: ZIP: 07090
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Foley, Shawn P.
: TELEPHONE: 908-654-5000
: TELEFAX: 908-654-7866
: TELEX: 139-125
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 74 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-463-224-3

Query Match 48.0%; Score 48; DB 1; Length 74;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RAARISIGPR 10
|iiiiiii|

```
Db      37 RAARISLGR 46

RESULT 2
US-08-463-377-3
; Sequence 3, Application US/08463377
; Patent No. 5837499
; GENERAL INFORMATION:
; APPLICANT: van Oostrum, Jan
; APPLICANT: Boyar, William C.
; APPLICANT: Galakatos, Nicholas G.
; APPLICANT: Schmitz, Albert
; APPLICANT: van Heeke, Gino
; TITLE OF INVENTION: C5a Receptor Antagonists Having
; SUBSTITUENT: Substantially No. 5837499Agonist Activity
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
; STREET: 600 South Avenue West
; CITY: Westfield
; STATE: NJ
; COUNTRY: USA
; ZIP: 07090
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,377
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Foley, Shawn P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-654-5000
; TELEFAX: 908-654-7866
; TELEX: 139-125
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-463-377-3

Query Match      48.0%; Score 48; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 RAARISLGR 10
      1111111111
Db      37 RAARISLGR 46

RESULT 3
US-09-246-500B-10
; Sequence 10, Application US/09246500B
; Patent No. 6235494
; GENERAL INFORMATION:
; APPLICANT: Hugli, Tony E.
; TITLE OF INVENTION: Substrates for Assessing Mannan-Binding
; PROTEIN: Protein-Associated Serine Protease Activity and Methods
; TITLE OF INVENTION: Using the Substrates
; FILE REFERENCE: 24730-2204
; CURRENT APPLICATION NUMBER: US/09/246,500B
; CURRENT FILING DATE: 1999-02-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 74
; TYPE: PRT

US-09-246-500B-10
; ORGANISM: Human C5a Anaphylatoxin
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: (64)...(66)
US-09-246-500B-10

Query Match      48.0%; Score 48; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 RAARISLGR 10
      1111111111
Db      37 RAARISLGR 46

RESULT 4
US-08-299-285-7
; Sequence 7, Application US/08299285
; Patent No. 5696230
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; C-TERM: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
US-08-299-285-7

Query Match      42.0%; Score 42; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      12 YSPKPMPLXR 21
      1111111111
Db      1 YSPKPMPLAR 10
```

RESULT 5
US-08-299-285-10
; Sequence 10, Application US/08299285
; Patent No. 5696230
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: "D-Alanine at position 9"
US-08-299-285-10

Query Match 42.0%; Score 42; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 12 YSPKPMPLXR 21
Db 1 YSPKPMPLAR 10

RESULT 6
US-08-299-285-24
; Sequence 24, Application US/08299285
; Patent No. 5696230
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 9
; OTHER INFORMATION: "D-Alanine at position 9"
US-08-299-285-24

Query Match 42.0%; Score 42; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 12 YSPKPMPLXR 21
Db 1 YSPKPMPLAR 10

RESULT 7
US-08-985-126-7
; Sequence 7, Application US/08985126
; Patent No. 5942599
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/985,126
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION NUMBER: US/08/299,285
;; FILING DATE: 31-AUG-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Reed, Janet E.
;; REGISTRATION NUMBER: 36,252
;; REFERENCE/DOCKET NUMBER: 63075
;; TELEPHONE: (215) 563-4100
;; TELEFAX: (215) 563-4044
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: No. 5942599 Relevant
;; TOPOLOGY: No. 5942599 Relevant
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: C-terminal
;; US-08-985-126-7

Query Match 42.0%; Score 42; DB 2; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 YSFKPMLXR 21
Db 1 YSFKDMLAR 10

RESULT 8
US-08-985-126-10
; Sequence 10, Application US/08985126
; Patent No. 5942599
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELEPHONE: (215) 563-4100

;; TELEFAX: (215) 563-4044
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: No. 5942599 Relevant
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: C-terminal
;; FEATURE:
;; NAME/KEY: Modified-site
;; LOCATION: 9
;; OTHER INFORMATION: "D-Alanine at position 9"
;; US-08-985-126-10

Query Match 42.0%; Score 42; DB 2; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 YSFKPMLXR 21
Db 1 YSFKDMLAR 10

RESULT 9
US-08-985-126-24
; Sequence 24, Application US/08985126
; Patent No. 5942599
; GENERAL INFORMATION:
; APPLICANT: Sanderson, Sam D.
; APPLICANT: Sherman, Simon A.
; APPLICANT: Kirnarsky, Leonid
; APPLICANT: Taylor, Stephen M.
; TITLE OF INVENTION: High Affinity Response-Selective
; C-Terminal Analogs of C5a Anaphylatoxin
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: US
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,126
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/299,285
; FILING DATE: 31-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; REFERENCE/DOCKET NUMBER: 63075
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
;; INFORMATION FOR SEQ ID NO: 24:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: No. 5942599 Relevant
;; TOPOLOGY: No. 5942599 Relevant
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO

; ANTI-SENSE: NO
 ; FRAGMENT TYPE: C-terminal
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 9
 ; OTHER INFORMATION: "D-Alanine at position 9"
 US-08-985-126-24

Query Match 42.0%; Score 42; DB 2; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 YSRKPMPLXR 21
 Db 1 YSRKPMQLAR 10

RESULT 10
 PCT-US95-11126-7
 ; Sequence 7, Application PC/TUS9511126
 ; GENERAL INFORMATION:
 ; APPLICANT: Sanderson, Sam D.
 ; APPLICANT: Sherman, Simon A.
 ; APPLICANT: Kirnarsky, Leonid
 ; APPLICANT: Taylor, Stephen M.
 ; TITLE OF INVENTION: High Affinity Response-Selective
 ; C-Terminal Analogs of C5a Anaphylatoxin
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
 ; STREET: 1601 Market Street Suite 720
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: US
 ; ZIP: 19103-2307
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/11126
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/299,285
 ; FILING DATE: 31-AUG-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Reed, Janet E.
 ; REGISTRATION NUMBER: 36,252
 ; REFERENCE/DOCKET NUMBER: 63075
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 563-4100
 ; TELEFAX: (215) 563-4044
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 10 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: C-terminal
 ; OTHER INFORMATION: /product= "D-Alanine at position 9"
 PCT-US95-11126-7

Query Match 42.0%; Score 42; DB 5; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 YSRKPMPLXR 21

Db 1 YSRKDMPLAR 10

RESULT 11
 PCT-US95-11126-10
 ; Sequence 10, Application PC/TUS9511126
 ; GENERAL INFORMATION:
 ; APPLICANT: Sanderson, Sam D.
 ; APPLICANT: Sherman, Simon A.
 ; APPLICANT: Kirnarsky, Leonid
 ; APPLICANT: Taylor, Stephen M.
 ; TITLE OF INVENTION: High Affinity Response-Selective
 ; C-Terminal Analogs of C5a Anaphylatoxin
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
 ; STREET: 1601 Market Street Suite 720
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: US
 ; ZIP: 19103-2307
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/11126
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/299,285
 ; FILING DATE: 31-AUG-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Reed, Janet E.
 ; REGISTRATION NUMBER: 36,252
 ; REFERENCE/DOCKET NUMBER: 63075
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 563-4100
 ; TELEFAX: (215) 563-4044
 ; INFORMATION FOR SEQ ID NO: 10:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 10 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: peptide
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; FRAGMENT TYPE: C-terminal
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 9
 ; OTHER INFORMATION: /product= "D-Alanine at position 9"
 PCT-US95-11126-10

Query Match 42.0%; Score 42; DB 5; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.22;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 YSRKPMPLXR 21
 Db 1 YSRKDMPLAR 10

RESULT 12
 PCT-US95-11126-24
 ; Sequence 24, Application PC/TUS9511126
 ; GENERAL INFORMATION:
 ; APPLICANT: Sanderson, Sam D.
 ; APPLICANT: Sherman, Simon A.

APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER-READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /product= "D-Alanine at position 9"
PCT-US95-11126-24

Query Match 42.0%; Score 42; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.22;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 YSPKPMPLXR 21
|||||
Db 1 YSPKPMOLAR 10

RESULT 13
US-08-299-285-5
Sequence 5, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia

STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER-READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-299-285-5
Query Match 41.0%; Score 41; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.33;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 12 YSPKPMPLXR 21
|||||
Db 1 YSPKPMPLGR 10
RESULT 14
US-08-299-285-13
Sequence 13, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER-READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 563-4100
 TELEFAX: (215) 563-4044
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: not relevant
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: C-terminal
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 9
 OTHER INFORMATION: "D-Phenylalanine at
 position 9"
 US-08-299-285-13

Query Match 41.0%; Score 41; DB 1; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.33;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 12 YSFKPMLXR 21
 Db 1 YSFKDMLPLR 10

RESULT 15
 US-08-985-126-5
 Sequence 5, Application US/08985126
 Patent No. 5942599
 GENERAL INFORMATION:
 APPLICANT: Sanderson, Sam D.
 APPLICANT: Sherman, Simon A.
 APPLICANT: Kirnarsky, Leonid
 TITLE OF INVENTION: High Affinity Response-Selective
 TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dann, Dorfman, Herrell and Skillman
 STREET: 1601 Market Street Suite 720
 CITY: Philadelphia
 STATE: PA
 COUNTRY: US
 ZIP: 19103-2307
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/985,126
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/299,285
 FILING DATE: 31-AUG-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Reed, Janet E.
 REGISTRATION NUMBER: 36,252
 REFERENCE/DOCKET NUMBER: 63075
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 563-4100
 TELEFAX: (215) 563-4044
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10 amino acids
 TYPE: amino acid
 STRANDEDNESS: NO. 5942599 Relevant

TOPOLOGY: No. 5942599 Relevant
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: C-terminal
 US-08-985-126-5

Query Match 41.0%; Score 41; DB 2; Length 10;
 Best Local Similarity 80.0%; Pred. No. 0.33;
 Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 12 YSFKPMLXR 21
 Db 1 YSFKDMLPLR 10

Search completed: February 27, 2002, 11:36:03
 Job time: 142 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:14 ; Search time 303.5 Seconds
(without alignments)
4.881 Million cell updates/sec

Title: US-09-446-109a-6
Perfect score: 99
Sequence: 1 KYKHSVVKKXSFKPMPLXR 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	54	54.5	12	18	AAW27254
2	53	53.5	19	18	AAW27251
3	50	50.5	10	18	AAW27250
4	50	50.5	18	18	AAW27256
5	47	47.5	9	22	AAW74059
6	47	47.5	15	22	AAW74086
7	47	47.5	16	22	AAW74085
8	47	47.5	17	22	AAW74084
9	47	47.5	18	22	AAW74083
10	47	47.5	19	22	AAW74082
11	47	47.5	20	22	AAW74054

12	47	47.5	20	22	AAW74087	C-terminal truncat
13	47	47.5	20	22	AAW74088	C-terminal truncat
14	47	47.5	20	22	AAW74089	C-terminal truncat
15	47	47.5	20	22	AAW74091	C-terminal truncat
16	47	47.5	20	22	AAW74092	C-terminal truncat
17	47	47.5	20	22	AAW74096	C-terminal truncat
18	47	47.5	64	18	AAW07784	Human complement C
19	47	47.5	65	18	AAW07785	Human complement C
20	47	47.5	66	18	AAW07786	Human complement C
21	47	47.5	67	18	AAW07787	Human complement C
22	47	47.5	68	18	AAW07788	Human complement C
23	47	47.5	69	18	AAW07789	Human complement C
24	47	47.5	70	18	AAW07790	Human complement C
25	47	47.5	71	18	AAW07782	Human complement C
26	47	47.5	71	18	AAW07783	Human complement C
27	47	47.5	71	18	AAW07804	Human complement C
28	47	47.5	71	18	AAW07791	Human complement C
29	47	47.5	72	18	AAW07792	Human complement C
30	47	47.5	74	8	AAW71666	Human anaphylatoxin
31	47	47.5	74	16	AAW75497	Human C5a protein.
32	47	47.5	74	22	AAE05454	Human C5a anaphyla
33	47	47.5	74	22	AAW74053	Human C5a. Homo s
34	47	47.5	74	22	AAW74119	Variant human C5a.
35	47	47.5	77	22	AAE05458	Mouse C5a anaphyla
36	47	47.5	1676	16	AAW77604	Pro-C5 polypeptide
37	46.5	47.0	1279	19	AAW98761	H. pylori GPO 109
38	44	44.4	9	22	AAW74063	C-terminal truncat
39	44	44.4	20	22	AAW74090	C-terminal truncat
40	43	43.4	9	22	AAW74064	C-terminal truncat
41	43	43.4	9	22	AAW74066	C-terminal truncat
42	43	43.4	20	22	AAW74093	C-terminal truncat
43	43	43.4	20	22	AAW74095	C-terminal truncat
44	42	42.4	10	17	AAW94487	C5a anaphylatoxin
45	42	42.4	10	17	AAW94470	C5a anaphylatoxin

ALIGNMENTS

RESULT	1					
AAW27254						
ID	AAW27254	standard; peptide; 12 AA.				
XX	XX	AAW27254;				
AC	AAW27254;					
XX	XX	23-DEC-1997 (first entry)				
DT	DT	Molecular adjuvant formula.				
XX	XX	Molecular adjuvant; Immune response; immunogen; binding affinity;				
DE	DE	antigen presenting cell; APC; viral pathogen; anti tumour response;				
XX	XX	antibody; immunodiagnostic; immunotherapeutic; agonist.				
KW	KW	Synthetic.				
XX	XX	Key				
FT	FT	Modified-site 1	Location/Qualifiers			
FT	FT	Modified-site 2	/note= "Serum amyloid A is attached to the N-terminal"			
FT	FT	/Label= OTHER				
FT	FT	/note= "Aminohexanoic acid which is a linear aliphatic spacer moiety"				
FT	FT	/note= "D-form residue"				
FT	FT	Misc-difference 11				
FT	FT					
XX	XX	WO9714426-A1.				
XX	XX	24-APR-1997.				
XX	XX	18-OCT-1996;	96WO-US16825.			
XX	XX	20-OCT-1995;	95US-0005727.			

PA (UYNE-) UNIV NEBRASKA.
 XX Hollingsworth MA, Sanderson SD, Tempero RA;
 XX WPI; 1997-244854/22.

XX Molecular adjuvants for enhancing immune responses - comprise an
 PT immunogen linked to a ligand having binding affinity for a
 PT determinant of an antigen presenting cell

XX Claim 15: Page 53: 61pp; English.

XX A novel molecular adjuvant has been developed for enhancing an immune
 CC response to an immunogen. The molecular adjuvant comprises a targeting
 CC ligand, having binding affinity for a characteristic determinant of an
 CC antigen presenting cell (APC), the targeting ligand being functionally
 CC linked to the immunogen, and so binding of the molecular adjuvant to
 CC the APC determinant activates the APC, effecting delivery of the
 CC immunogen to an antigen presenting pathway of the APC. The present
 CC sequence represents a specifically claimed formula for a molecular
 CC adjuvant, where the immunogen is serum amyloid A. The molecular adjuvant
 CC can be used to produce an immune response for protecting against viral
 CC and other pathogens or to produce anti-tumour responses. It can also be
 CC used for the production of antibodies for use as immunodiagnostic and
 CC immunotherapeutic agents. The molecular adjuvant is readily taken up and
 CC processed by APCs to provide APC-mediated immune responses. It can
 CC provide for the production of antibodies against a weakly-antigenic or
 CC non-antigenic substances.

XX Sequence 12 AA;

Query Match 54.5%; Score 54; DB 18; Length 12;
 Best Local Similarity 91.7%; Pred. No. 0.0066;
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 KYXSKPMLPLXR 20
 II II II II II II II I
 Db 1 kxysfkpmplar 12

RESULT 2
 AAW27251
 ID AAW27251 standard; peptide: 19 AA.

AC AAW27251;
 XX 23-DEC-1997 (first entry)
 DT Molecular adjuvant targeting ligand and immunogen.

DE Molecular adjuvant; immune response; immunogen; binding affinity;
 XX antigen presenting cell; APC; viral pathogen; anti tumour response;
 KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX Synthetic.

XX Key Location/Qualifiers
 FH Misc-difference 18
 FT /note= "D-form residue"

XX WO9714426-A1.

XX 24-APR-1997.

XX 18-OCT-1996; 96WO-US16825.

XX 20-OCT-1995; 95US-0005727.

XX (UYNE-) UNIV NEBRASKA.

XX Hollingsworth MA, Sanderson SD, Tempero RA;

DR WPI; 1997-244854/22.

XX Molecular adjuvants for enhancing immune responses - comprise an
 PT immunogen linked to a ligand having binding affinity for a
 PT determinant of an antigen presenting cell

XX Claim 6: Page 52: 61pp; English.

XX A novel molecular adjuvant has been developed for enhancing an immune
 CC response to an immunogen. The molecular adjuvant comprises a targeting
 CC ligand, having binding affinity for a characteristic determinant of an
 CC antigen presenting cell (APC), the targeting ligand being functionally
 CC linked to the immunogen, and so binding of the molecular adjuvant to
 CC the APC determinant activates the APC, effecting delivery of the
 CC immunogen to an antigen presenting pathway of the APC. The present
 CC sequence represents a specifically claimed targeting ligand and
 CC immunogen. The molecular adjuvant can be used to produce an immune
 CC response for protecting against viral and other pathogens or to
 CC produce anti-tumour responses. It can also be used for the production
 CC of antibodies for use as immunodiagnostic and immunotherapeutic
 CC agents. The molecular adjuvant is readily taken up and processed
 CC by APCs to provide APC-mediated immune responses. It can provide
 CC for the production of antibodies against a weakly-antigenic or
 CC non-antigenic substances.

XX Sequence 19 AA;

Query Match 53.5%; Score 53; DB 18; Length 19;
 Best Local Similarity 57.9%; Pred. No. 0.016;
 Matches 11; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 2 YKHSVVKXYSFKPMLPLXR 20
 II II II II II II II I
 Db 1 ykqggfllgylstkpmplar 19

RESULT 3
 AAW27250
 ID AAW27250 standard; peptide: 10 AA.

XX AAW27250;

DT 23-DEC-1997 (first entry)

XX Molecular adjuvant targeting ligand.

XX Molecular adjuvant; immune response; immunogen; binding affinity;
 KW antigen presenting cell; APC; viral pathogen; anti tumour response;
 KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX Synthetic.

XX Key Location/Qualifiers
 FH Misc-difference 9
 FT /note= "D-form residue"

XX WO9714426-A1.

XX 24-APR-1997.

XX 18-OCT-1996; 96WO-US16825.

XX 20-OCT-1995; 95US-0005727.

XX (UYNE-) UNIV NEBRASKA.

XX Hollingsworth MA, Sanderson SD, Tempero RA;

XX WPI; 1997-244854/22.

XX Molecular adjuvants for enhancing immune responses - comprise an
 PT immunogen linked to a ligand having binding affinity for a

PT determinant of an antigen presenting cell

PS Claim 5; page 51; 61pp; English.

A novel molecular adjuvant has been developed for enhancing an immune response to an immunogen. The molecular adjuvant comprises a targeting ligand, having binding affinity for a characteristic determinant of an antigen presenting cell (APC), the targeting ligand being functionally linked to the immunogen, and so binding of the molecular adjuvant to the APC determinant activates the APC, effecting delivery of the immunogen to an antigen presenting pathway of the APC. The present sequence represents a specifically claimed targeting ligand, which is a C5a C-terminal decapeptide agonist. The molecular adjuvant can be used to produce an immune response for protecting against viral and other pathogens or to produce anti-tumour responses. It can also be used for the production of antibodies for use as immunodiagnostic and immunotherapeutic agents. The molecular adjuvant is readily taken up and processed by APCs to provide APC-mediated immune responses. It can provide for the production of antibodies against a weakly-antigenic or non-antigenic substances.

XX	Sequence	10 AA:
S0		

Query Match 50.5%; Score 50; DB 18; Length 10;
 Best Local Similarity 90.0%; Pred. No. 0.025;
 Matches 9; Conservative 0; Mismatches 1; Indels

Qy 11 YSFKPMLXR 20
1111111111
Db 1 ysfkpmplxr 10

RESULT

AAW27236
ID AAW27256 standard: peptide: 18 AA.

XX AAW27256:

XX
DT 23-DEC-1997 (first entry)

XX Molecular adjuvant targeting ligand and immunogen.

XX Molecular adjuvant; immune response; immunogen; binding affinity;
KW antigen presenting cell; APC; viral pathogen; anti tumour response;
KW antibody; immunodiagnostic; immunotherapeutic; agonist.

XX
OS
Synthetic

XX	Key	Location/Qualifiers
FH		

FT	Misc-difference	9	/note= "D-form residue"
FT			

XX
PN W09714426-A1XX
PD
24-APR-1997

XX 18-OCT-1996 96W0-11S16825

XX
20-00T-1995. 95115-00005727XX
DA (IYNE-) INTV NEBRASKA

XX PI Hollingsworth MA: Sanderson SD: Tempero RA:

XX
DP
WPT: 1997-244854/22

XX	
PT	Molecular adjuvants for enhancing immune responses - comprise an immunogen linked to a ligand having binding affinity for a determinant of an antigen presenting cell
PT	

Example 1: Page 31; 61pp; English.

A novel molecular adjuvant has been developed for enhancing an immune response to an immunogen. The molecular adjuvant comprises a targeting ligand, having binding affinity for a characteristic determinant of an antigen presenting cell (APC), the targeting ligand being functionally linked to the immunogen, and so binding of the molecular adjuvant to the APC determinant activates the APC, effecting delivery of the immunogen to an antigen presenting pathway of the APC. The present sequence represents a targeting ligand and immunogen, in which the juxta-membrane epitope of human mucin-1 is positioned toward the carboxyl terminus and the C5a analogue is positioned toward the amino terminus. The molecular adjuvant can be used to produce an immune response for protecting against viral and other pathogens or to produce anti-tumour responses. It can also be used for the production of antibodies for use as immunodiagnostic and immunotherapeutic agents. The molecular adjuvant is readily taken up and processed by APCs to provide APC-mediated immune responses. It can provide for the production of antibodies against a weakly-antigenic or non-antigenic substances.

Sequence 18 AA:

Query Match 50.5%; Score 50; DB 18; Length 18;
Best Local Similarity 90.0%; Pred. No. 0.05;
Matches 9: Conservative 0; Mismatches 1: Indels

Qy . 11 YSEKPMPLXR 20
+ + + + +
Db 1 ysfkpmplxr 10

RESULT 5

AAB/4059
ID AAB74059 standard: peptide: 9 AA.

AA AAB74059:

DT 16-MAY-2001 (first entry)

DE Human C5a peptide fragment #4.

XX Human; C5a; complement; antibody; bacterial infection; sinusitis;
KW meningitis; respiratory; gastrointestinal; urinary tract infection;
KW wound; anaphylatoxin; sepsis.

XX US Homo sapiens.

PN WO200115771-A1

XX PD 08-MAR-2001.

XX 31-AUG-2000: 2000WO-US24219.

XX 31-AUG-1990. 99US-0387671. PR

XX
PA
(JINMT) JIN TV MICHIGAN.

XX
PT
Ward PA. Huber-Lang M. Sarma V:

XX
DR WPT: 2001-226665/23.XX
DR N-PSDB; AAF/5796.

XX Compositions for treating blood-borne and toxin mediated diseases and
PT treatment of sepsis in humans and other animals comprising anti-C5a
PT antibodies generated against C-terminal truncated C5a peptides -

XX
PS
Claim 8: Page 26: English.

The present sequence is a peptide fragment of human complement component C5a (the full-length sequence is given in A874053). The present invention relates to an antibody specific for the present sequence. The C5a-antibody can be used in a therapeutic composition, which is useful for treating a subject suffering from bacterial infection, e.g.

CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract
 CC infections or infections in wounds. In addition, the C5a antibody can
 CC be used for treating sepsis. C5a is also known as anaphylatoxin.

XX SQ Sequence 9 AA;

Query Match 47.5%; Score 47; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYKHSVVKK 9
 |||||
 Db 1 kykhsvvkk 9

RESULT 6
 AAB74086
 ID AAB74086 standard; Peptide; 15 AA.

XX AC AAB74086;

DT 16-MAY-2001 (first entry)

DE C-terminal truncated C5a peptide #24.

CC C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.

XX OS Unidentified.

PN WO200115731-A1.

XX PD 08-MAR-2001.

PF 31-AUG-2000; 2000WO-US24219.

PR 31-AUG-1999; 99US-0387671.

XX PA (UNMI) UNIV MICHIGAN.

XX PI Ward PA, Huber-Lang M, Sarma V;

XX DR WPI; 2001-226665/23.

XX PT Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides .

XX PS Disclosure; Page 29; 84pp: English.

XX CC The present sequence is a C-terminal truncated C5a peptide fragment. The
 CC present invention relates to an antibody specific for the present
 CC sequence. The C5a-antibody can be used in a therapeutic composition,
 CC which is useful for treating a subject suffering from bacterial
 CC infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
 CC urinary tract infections or infections in wounds. In addition, the C5a
 CC antibody can be used for treating sepsis. C5a is also known as
 CC anaphylatoxin.

XX SQ Sequence 15 AA;

Query Match 47.5%; Score 47; DB 22; Length 15;
 Best Local Similarity 100.0%; Pred. No. 0.13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYKHSVVKK 9
 |||||
 Db 7 kykhsvvkk 15

RESULT 7
 AAB74085

ID AAB74085 standard; Peptide; 16 AA.

XX AC AAB74085;

DT 16-MAY-2001 (first entry)

DE C-terminal truncated C5a peptide #23.

CC C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.

XX OS Unidentified.

XX PN WO200115731-A1.

XX PD 08-MAR-2001.

PF 31-AUG-2000; 2000WO-US24219.

PR 31-AUG-1999; 99US-0387671.

XX PA (UNMI) UNIV MICHIGAN.

XX PI Ward PA, Huber-Lang M, Sarma V;

XX DR WPI; 2001-226665/23.

XX PT Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides .

XX PS Disclosure; Page 29; 84pp: English.

XX CC The present sequence is a C-terminal truncated C5a peptide fragment. The
 CC present invention relates to an antibody specific for the present
 CC sequence. The C5a-antibody can be used in a therapeutic composition,
 CC which is useful for treating a subject suffering from bacterial
 CC infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
 CC urinary tract infections or infections in wounds. In addition, the C5a
 CC antibody can be used for treating sepsis. C5a is also known as
 CC anaphylatoxin.

XX SQ Sequence 16 AA;

Query Match 47.5%; Score 47; DB 22; Length 16;
 Best Local Similarity 100.0%; Pred. No. 0.14;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYKHSVVKK 9
 |||||
 Db 8 kykhsvvkk 16

RESULT 8
 AAB74084

ID AAB74084 standard; Peptide; 17 AA.

XX AC AAB74084;

DT 16-MAY-2001 (first entry)

DE C-terminal truncated C5a peptide #22.

CC C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.

XX OS Unidentified.

PN WO200115731-A1.
 XX 08-MAR-2001.
 XX 31-AUG-2000; 2000WO-US24219.
 XX 31-AUG-1999; 99US-0387671.
 XX (UNMI) UNIV MICHIGAN.
 XX Ward PA, Huber-Lang M, Sarma V;
 XX WPI; 2001-226665/23.
 XX Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides
 XX
 PS Disclosure; Page 29; 84pp; English.
 XX The present sequence is a C-terminal truncated C5a peptide fragment. The
 CC present invention relates to an antibody specific for the present
 CC sequence. The C5a-antibody can be used in a therapeutic composition,
 CC which is useful for treating a subject suffering from bacterial
 CC infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
 CC urinary tract infections or infections in wounds. In addition, the C5a
 CC antibody can be used for treating sepsis. C5a is also known as
 CC anaphylatoxin.
 XX Sequence 17 AA;
 SQ

Query Match 47.5%; Score 47; DB 22; Length 17;
 Best Local Similarity 100.0%; Pred. No. 0.15;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYKHSVVKK 9
 Db 9 kykhsvvkk 17
 |||||

RESULT 9
 AAB74083
 ID AAB74083 standard; Peptide; 18 AA.
 XX
 AC AAB74083;
 XX
 DT 16-MAY-2001 (first entry)
 XX
 DE C-terminal truncated C5a peptide #21.
 XX
 KW C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.
 XX
 OS Unidentified.
 XX
 PN WO200115731-A1.
 XX
 PD 08-MAR-2001.
 XX
 PF 31-AUG-2000; 2000WO-US24219.
 XX
 PR 31-AUG-1999; 99US-0387671.
 XX
 PA (UNMI) UNIV MICHIGAN.
 XX
 PI Ward PA, Huber-Lang M, Sarma V;
 XX
 WPI; 2001-226665/23.
 XX
 XX Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a

PT antibodies generated against C-terminal truncated C5a peptides
 XX
 PS Disclosure; Page 29; 84pp; English.
 XX The present sequence is a C-terminal truncated C5a peptide fragment. The
 CC present invention relates to an antibody specific for the present
 CC sequence. The C5a-antibody can be used in a therapeutic composition,
 CC which is useful for treating a subject suffering from bacterial
 CC infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
 CC urinary tract infections or infections in wounds. In addition, the C5a
 CC antibody can be used for treating sepsis. C5a is also known as
 CC anaphylatoxin.
 XX Sequence 18 AA;
 SQ

Query Match 47.5%; Score 47; DB 22; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.16;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYKHSVVKK 9
 Db 10 kykhsvvkk 18
 |||||

RESULT 10
 AAB74082
 ID AAB74082 standard; Peptide; 19 AA.
 XX
 AC AAB74082;
 XX
 DT 16-MAY-2001 (first entry)
 XX
 DE C-terminal truncated C5a peptide #20.
 XX
 KW C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.
 XX
 OS Unidentified.
 XX
 PN WO200115731-A1.
 XX
 PD 08-MAR-2001.
 XX
 PF 31-AUG-2000; 2000WO-US24219.
 XX
 PR 31-AUG-1999; 99US-0387671.
 XX
 PA (UNMI) UNIV MICHIGAN.
 XX
 PI Ward PA, Huber-Lang M, Sarma V;
 XX
 WPI; 2001-226665/23.
 XX
 XX Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides
 XX
 PS Disclosure; Page 29; 84pp; English.
 XX The present sequence is a C-terminal truncated C5a peptide fragment. The
 CC present invention relates to an antibody specific for the present
 CC sequence. The C5a-antibody can be used in a therapeutic composition,
 CC which is useful for treating a subject suffering from bacterial
 CC infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
 CC urinary tract infections or infections in wounds. In addition, the C5a
 CC antibody can be used for treating sepsis. C5a is also known as
 CC anaphylatoxin.
 XX Sequence 19 AA;
 SQ

Query Match 47.5%; Score 47; DB 22; Length 19;
 Best Local Similarity 100.0%; Pred. No. 0.17; Mismatches 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYKHSVVKK 9
 Db 11 kykhsvvkk 19

RESULT 11
 AAB74054
 ID AAB74054 standard; Peptide: 20 AA.
 XX
 AC AAB74054;
 DT 16-MAY-2001 (first entry)
 DE Human C5a peptide fragment #1.
 XX Human; C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.
 XX Homo sapiens.
 OS
 XX
 XX WO200115731-A1.
 PN
 XX
 XX 08-MAR-2001.
 PD
 XX
 XX 31-AUG-2000; 2000WO-US24219.
 PF
 XX
 XX 31-AUG-1999; 99US-0387671.
 PR
 XX
 XX (UNMI) UNIV MICHIGAN.
 PA
 XX Ward PA, Huber-Lang M, Sarma V;
 PI
 XX WPI: 2001-226665/23.
 DR
 XX
 XX Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides -
 XX
 PS Claim 8; Page 26; 84pp; English.
 XX
 CC The present sequence is a peptide fragment of human complement component
 CC C5a (the full-length sequence is given in AAB74053). The present
 CC invention relates to an antibody specific for the present sequence. The
 CC C5a-antibody can be used in a therapeutic composition, which is useful
 CC for treating a subject suffering from bacterial infection, e.g.
 CC sinusitis, meningitis, respiratory, gastrointestinal or urinary tract
 CC infections or infections in wounds. In addition, the C5a antibody can
 CC be used for treating sepsis. C5a is also known as anaphylatoxin.
 XX
 XX Sequence 20 AA;

Query Match 47.5%; Score 47; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.18; Mismatches 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYKHSVVKK 9
 Db 12 kykhsvvkk 20

RESULT 12
 AAB74087
 ID AAB74087 standard; Peptide: 20 AA.
 XX
 AC AAB74087;
 DT 16-MAY-2001 (first entry)
 DE C-terminal truncated C5a peptide #26.
 XX C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.
 XX
 OS Unidentified.
 XX
 XX WO200115731-A1.
 PN
 XX
 XX 08-MAR-2001.
 PD
 XX
 XX 31-AUG-2000; 2000WO-US24219.
 PF
 XX

DT 16-MAY-2001 (first entry)
 XX
 DE C-terminal truncated C5a peptide #25.
 XX
 KW C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.
 XX
 OS Unidentified.
 XX
 XX WO200115731-A1.
 PN
 XX
 XX 08-MAR-2001.
 PD
 XX
 XX 31-AUG-2000; 2000WO-US24219.
 PF
 XX
 XX 31-AUG-1999; 99US-0387671.
 PR
 XX
 XX (UNMI) UNIV MICHIGAN.
 PA
 XX Ward PA, Huber-Lang M, Sarma V;
 PI
 XX WPI: 2001-226665/23.
 DR
 XX
 XX Compositions for treating blood-borne and toxin mediated diseases and
 PT treatment of sepsis in humans and other animals comprises anti-C5a
 PT antibodies generated against C-terminal truncated C5a peptides -
 XX
 PS Disclosure; Page 29; 84pp; English.
 XX
 CC The present sequence is a C-terminal truncated C5a peptide fragment. The
 CC present invention relates to an antibody specific for the present
 CC sequence. The C5a-antibody can be used in a therapeutic composition,
 CC which is useful for treating a subject suffering from bacterial
 CC infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
 CC urinary tract infections or infections in wounds. In addition, the C5a
 CC antibody can be used for treating sepsis. C5a is also known as
 CC anaphylatoxin.
 XX
 XX Sequence 20 AA;

Query Match 47.5%; Score 47; DB 22; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.18; Mismatches 0; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KYKHSVVKK 9
 Db 12 kykhsvvkk 20

RESULT 13
 AAB74088
 ID AAB74088 standard; Peptide: 20 AA.
 XX
 AC AAB74088;
 DT 16-MAY-2001 (first entry)
 DE C-terminal truncated C5a peptide #26.
 XX C5a; complement; antibody; bacterial infection; sinusitis;
 KW meningitis; respiratory; gastrointestinal; urinary tract infection;
 KW wound; anaphylatoxin; sepsis.
 XX
 OS Unidentified.
 XX
 XX WO200115731-A1.
 PN
 XX
 XX 08-MAR-2001.
 PD
 XX
 XX 31-AUG-2000; 2000WO-US24219.
 PF
 XX

```

PR 31-AUG-1999; 99US-0387671.
XX (UNMI ) UNIV MICHIGAN.
XX Ward PA, Huber-Lang M, Sarma V;
XX WPI; 2001-226665/23.
XX Compositions for treating blood-borne and toxin mediated diseases and
PT treatment of sepsis in humans and other animals comprises anti-C5a
PT antibodies generated against C-terminal truncated C5a peptides -
XX Disclosure; Page 29; 84pp; English.
XX The present sequence is a C-terminal truncated C5a peptide fragment. The
CC present invention relates to an antibody specific for the present
CC sequence. The C5a-antibody can be used in a therapeutic composition,
CC which is useful for treating a subject suffering from bacterial
CC infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
CC urinary tract infections or infections in wounds. In addition, the C5a
CC antibody can be used for treating sepsis. C5a is also known as
CC anaphylatoxin.
XX Sequence 20 AA:
PT Query Match 47.5%; Score 47; DB 22; Length 20;
PT Best Local Similarity 100.0%; Pred. No. 0.18;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 KYKHSVVKK 9
DB 12 kykhsvvkk 20
XXXXXXXXXX

RESULT 14
AAB74089
ID AAB74089 standard; Peptide; 20 AA.
XX
XX AAB74089:
XX 16-MAY-2001 (first entry)
XX C-terminal truncated C5a peptide #27.
XX C5a; complement; antibody; bacterial infection; sinusitis;
XX meningitis; respiratory; gastrointestinal; urinary tract infection;
XX wound; anaphylatoxin; sepsis.
XX Unidentified.
XX WO200115731-A1.
XX 08-MAR-2001.
XX 31-AUG-2000; 2000WO-US24219.
XX 31-AUG-1999; 99US-0387671.
XX (UNMI ) UNIV MICHIGAN.
XX Ward PA, Huber-Lang M, Sarma V;
XX WPI; 2001-226665/23.
XX Compositions for treating blood-borne and toxin mediated diseases and
XX treatment of sepsis in humans and other animals comprises anti-C5a
XX antibodies generated against C-terminal truncated C5a peptides -
XX Disclosure; Page 30; 84pp; English.
XX The present sequence is a C-terminal truncated C5a peptide fragment. The
XX present invention relates to an antibody specific for the present
XX sequence. The C5a-antibody can be used in a therapeutic composition,
XX which is useful for treating a subject suffering from bacterial
XX infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
XX urinary tract infections or infections in wounds. In addition, the C5a
XX antibody can be used for treating sepsis. C5a is also known as
XX anaphylatoxin.
XX Sequence 20 AA:
PT Query Match 47.5%; Score 47; DB 22; Length 20;
PT Best Local Similarity 100.0%; Pred. No. 0.18;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 KYKHSVVKK 9
DB 12 kykhsvvkk 20
XXXXXXXXXX

RESULT 15
AAB74091
ID AAB74091 standard; Peptide; 20 AA.
XX
XX AAB74091:
XX 16-MAY-2001 (first entry)
XX C-terminal truncated C5a peptide #29.
XX C5a; complement; antibody; bacterial infection; sinusitis;
XX meningitis; respiratory; gastrointestinal; urinary tract infection;
XX wound; anaphylatoxin; sepsis.
XX Unidentified.
XX WO200115731-A1.
XX 08-MAR-2001.
XX 31-AUG-2000; 2000WO-US24219.
XX 31-AUG-1999; 99US-0387671.
XX (UNMI ) UNIV MICHIGAN.
XX Ward PA, Huber-Lang M, Sarma V;
XX WPI; 2001-226665/23.
XX Compositions for treating blood-borne and toxin mediated diseases and
XX treatment of sepsis in humans and other animals comprises anti-C5a
XX antibodies generated against C-terminal truncated C5a peptides -
XX Disclosure; Page 30; 84pp; English.
XX The present sequence is a C-terminal truncated C5a peptide fragment. The
XX present invention relates to an antibody specific for the present
XX sequence. The C5a-antibody can be used in a therapeutic composition,
XX which is useful for treating a subject suffering from bacterial
XX infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
XX urinary tract infections or infections in wounds. In addition, the C5a
XX antibody can be used for treating sepsis. C5a is also known as
XX anaphylatoxin.
XX Sequence 20 AA:
PT Query Match 47.5%; Score 47; DB 22; Length 20;
PT Best Local Similarity 100.0%; Pred. No. 0.18;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 KYKHSVVKK 9
DB 12 kykhsvvkk 20
XXXXXXXXXX

RESULT 15
AAB74091
ID AAB74091 standard; Peptide; 20 AA.
XX
XX AAB74091:
XX 16-MAY-2001 (first entry)
XX C-terminal truncated C5a peptide #29.
XX C5a; complement; antibody; bacterial infection; sinusitis;
XX meningitis; respiratory; gastrointestinal; urinary tract infection;
XX wound; anaphylatoxin; sepsis.
XX Unidentified.
XX WO200115731-A1.
XX 08-MAR-2001.
XX 31-AUG-2000; 2000WO-US24219.
XX 31-AUG-1999; 99US-0387671.
XX (UNMI ) UNIV MICHIGAN.
XX Ward PA, Huber-Lang M, Sarma V;
XX WPI; 2001-226665/23.
XX Compositions for treating blood-borne and toxin mediated diseases and
XX treatment of sepsis in humans and other animals comprises anti-C5a
XX antibodies generated against C-terminal truncated C5a peptides -
XX Disclosure; Page 30; 84pp; English.
XX The present sequence is a C-terminal truncated C5a peptide fragment. The
XX present invention relates to an antibody specific for the present
XX sequence. The C5a-antibody can be used in a therapeutic composition,
XX which is useful for treating a subject suffering from bacterial
XX infection, e.g. sinusitis, meningitis, respiratory, gastrointestinal or
XX urinary tract infections or infections in wounds. In addition, the C5a
XX antibody can be used for treating sepsis. C5a is also known as
XX anaphylatoxin.
XX Sequence 20 AA:
PT Query Match 47.5%; Score 47; DB 22; Length 20;
PT Best Local Similarity 100.0%; Pred. No. 0.18;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 KYKHSVVKK 9
DB 12 kykhsvvkk 20
XXXXXXXXXX

```

Db 12 kykhsvkk 20

Search completed: February 27, 2002, 11:41:14
Job time: 453 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:16 ; Search time 145.23 Seconds
(without alignments)
10.490 Million cell updates/sec

Title: US-09-446-109A-6
 Perfect score: 99
 Sequence: 1 KYKHSVVKKXXYSFKPMP LXR 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	51	51.5	241	2	T17860	hypothetical prote	
2	48	48.5	2219	2	T27684	hypothetical prote	
3	47	47.5	1676	1	C5HU	complement C5 prec	
4	46.5	47.0	1279	2	E64709	type IIS restricti	
5	45	45.5	1170	2	T25043	hypothetical prote	
6	44	44.4	1313	2	G82887	hypothetical prote	
7	43	43.4	346	2	E72672	cysteine proteinas	
8	42	42.4	101	2	JC4792	cysteine proteinas	
9	42	42.4	123	2	JE0308	probable iron-sulf	
10	42	42.4	479	2	D81423	El protein - cotto	
11	42	42.4	602	1	W1WLR8	F819 protein - hum	
12	42	42.4	940	2	JE0291	peridinin-chloroph	
13	41	41.4	370	2	S60187	4-comarate-CoA li	
14	41	41.4	566	2	F85214	4-comarate--CoA li	
15	41	41.4	566	2	T05038	hypothetical prote	
16	41	41.4	1098	2	T28816	complement C5 prec	
17	41	41.4	1680	1	C5MS	hypothetical prote	
18	40	40.4	110	2	D70373	30k protein - pepp	
19	40	40.4	257	1	WMTNP3	hypothetical prote	
20	40	40.4	288	2	D83267	hypothetical prote	
21	40	40.4	413	2	H83833	probable membrane	
22	40	40.4	996	2	T25512	hypothetical prote	
23	40	40.4	4910	2	S64942	formyltetrahydrofo	
24	39	39.4	192	2	A84727	conserved hypothet	
25	39	39.4	283	2	C83105	hypothetical prote	
26	39	39.4	418	2	B72245	hypothetical prote	
27	39	39.4	464	2	T48339	typeIII site-speci	
28	39	39.4	468	2	F71422		
29	39	39.4	489	2	A81166		

ALIGNMENTS

```

RESULT      1
Tl7860
hypothetical protein A360R - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: Tl7860
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: Tl7860
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-241 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96728.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: A360R

```

Query Match	51.5%	Score 51;	DB 2;	Length 241;
Best Local Similarity	52.6%	Pred. No. 0.25;		
Matches 10: Conservative	2;	Mismatches	7;	Indels
				Gaps 0;

QY 2 YKHSVVKKXYSEKPMPLXR 20
: | | | | | | | | | |
PB 21 EKRSVVKKPIGKPLPARR 39

RESULT 2
T27684
hypothetical protein ZK1067.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27684
R:Thomas, K.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20404
A:Accession: T27684
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2219 <WIL>
A:Cross-references: EMBL:Z70038; PIDN:CAA93884.1; GSPDB:GN00020; CESP:ZK1067.2
A:Experimental source: clone ZK1067
C:Genetics:
A:Gene: CESP:ZK1067.2
A:Map position: 2
A:Introns: 139/3; 246/3; 487/3; 784/1; 960/2; 1067/3; 1113/3; 1171/3; 1321/2; 1383/3;

Query Match	48.5%	Score 48;	DB 2;	Length 2219;
Best Local Similarity	47.6%	Pred. NO. 8.9;		
Matches 10: Conservative	4;	Mismatches	5;	Indels 2; Gaps 1;

UDP-glucosyl trans
MPL protein - fis
E1 protein - deer
hypothetical prote
hypothetical prote
hypothetical serin
hypothetical prote
DNA-directed DNA p
probable immediat
hypothetical prote
hypothetical prote
neutral amino acid
ecdysteroid UDPglu
exinuclease ABC C
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote

```

Query Match          47.5%; Score 47; DB 1; Length 1676;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

QY    1 KYKHSVVKK  9
      | |||||
DB    689 KYKHSVVKK 697

RESULT      4
E64709
Type IIS restriction enzyme R and M protein - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C>Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C:Accession: E64709
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKee, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, A.; Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: E64709
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1279 <TOM>
A:Cross-references: GB:AE000649; GB:AE000511; NID:g2314687; PIDN:AAD08556.1; PID:g2314687

Query Match          47.0%; Score 46.5; DB 2; Length 1279;
Best Local Similarity 45.5%; Pred. No. 9;
Matches     10; Conservative      5; Mismatches      4; Indels      3; Gaps      1;

QY    2 YKHSVKKXYSF---KPMLXR 20
      ||:||| : |||| :
DB   369 YKNSVLKKHKDYQKEKPLLLK 390

RESULT      5
T25043
hypothetical protein T21B10.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T25043
R:Baynes, C.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z19973
A:Accession: T25043
A:Molecule type: DNA
A>Status: preliminary; translated from GB/EMBL/DDBBJ
A:Molecule type: DNA
A:Residues: 1-1170 <WIL>
A:Cross-references: EMBL:Z68318; PIDN:CAA92693.2; GSPDB:GN00020; CESP:T21B10.3
A:Experimental source: clone T21B10
C:Genetics:
A:Gene: CESP-T21B10.3
A:Map position: 2
A:Supernum: 11/3; 45/1; 140/1; 206/3; 244/2; 342/3; 696/2; 767/3; 863/3; 914/1; 1052/2
C:Superfamily: Caenorhabditis elegans hypothetical protein T21B10.3

```

hypothetical protein UU474 [imported] - Ureaplasma urealyticum

C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82887
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a minor
A:Reference number: A82870
A:Accession: G82887
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1313 <GLA>
A:Cross-references: GB:AE002143; GB:AF222894; NID:G6899457; PIDN:AAF30886.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UU474
A:Genetic code: SGC3

Query Match 44.4%; Score 44; DB 2; Length 1313;
Best Local Similarity 63.6%; Pred. No. 25;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 3 KHSVVVKXXYSF 13
IIIIIIIIII

Db 49 KHTLKKYYSF 59

RESULT 7

E72672
hypothetical protein APE0803 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 08-Sep-2000
C:Accession: E72672
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339
A:Accession: E72672
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <KAW>
A:Cross-references: DDBJ:AF0000060; NID:G5104188; PIDN:BAA79781.1; PID:dl043567; PID:G510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0803
C:Superfamily: Aeropyrum pernix hypothetical protein APE0803

Query Match 43.4%; Score 43; DB 2; Length 346;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 VVKXXYSFKPMP 17
IIIIIIIIII

Db 12 VVKGTSLKPLP 23

RESULT 8

JC4792

cysteine proteinase inhibitor Scb - common sunflower
C:Species: Helianthus annuus (common sunflower)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 07-May-1999
C:Accession: JC4792
R:Kouzuma, Y.; Kawano, K.; Kimura, M.; Yamasaki, N.; Kadowaki, T.; Yamamoto, K.
J. Biochem. 119, 1106-1113, 1996
A:Title: Purification, characterization, and sequencing of two cysteine proteinase inhib
A:Reference number: JC4791; MUID:96424994
A:Accession: JC4792
A:Molecule type: Protein
A:Residues: 1-101 <KOU>
A:Experimental source: seed

C:Comment: This inhibitor is involved in protection against exogenous enzymes resembl
C:Superfamily: cystatin; cystatin homology
C:Keywords: cysteine proteinase inhibitor; seed

Query Match 42.4%; Score 42; DB 2; Length 101;
Best Local Similarity 47.4%; Pred. No. 3.6;
Matches 9; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 YKHSVVVKXXYSFKPMP 20
IIIIIIIIII

Db 85 WKHS--KLLGFRKAPVVK 101

RESULT 9

JE0308
cysteine proteinase inhibitor - common sunflower
C:Species: Helianthus annuus (common sunflower)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C:Accession: JE0308
R:Doi-Kawano, K.; Kouzuma, Y.; Yamasaki, N.; Kimura, M.
J. Biochem. 124, 911-916, 1998
A:Title: Molecular cloning, functional expression, and mutagenesis of cDNA encoding
A:Reference number: JE0308; MUID:99011314
A:Accession: JE0308
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-123 <DOI>
C:Superfamily: cystatin; cystatin homology

Query Match 42.4%; Score 42; DB 2; Length 123;
Best Local Similarity 47.4%; Pred. No. 4.5;
Matches 9; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 YKHSVVVKXXYSFKPMP 20
IIIIIIIIII

Db 107 WKHS--KLLGFRKAPVVK 123

RESULT 10

DB1423
probable iron-sulfur protein Cj0074c [imported] - Campylobacter jejuni (strain NCTC 1
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: DB1423
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: AB1250; MUID:20150912
A:Accession: DB1423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-479 <PAR>
A:Cross-references: GB:AL139074; GB:AL111168; NID:G6967505; PIDN:CAB72560.1; PID:G696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0074c
C:Superfamily: conserved hypothetical iron-sulfur protein HP0138; ferredoxin 2[4Fe-4S

Query Match 42.4%; Score 42; DB 2; Length 479;
Best Local Similarity 58.3%; Pred. No. 19;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 VVKXXYSFKPMP 17
IIIIIIIIII

Db 449 VIKKWAYAFKELP 460

RESULT 11

W1WLRB

#keywords: chloroplast
 F:1-57/Domain: transit peptide (chloroplast) #status predicted <TNP>
 F:58-370/Product: peridinin-chlorophyll a-binding protein #status predicted <MAT>

QY 4 HSVVKKXYSFKPMLXR 20
:|::|: :| | | :
Db 186 YSIMKESFGVPPKPLIK 202

Search completed: February 27, 2002, 11:45:17
Job time: 696 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:43 ; Search time 78.39 seconds
(without alignments)
9.354 Million cell updates/sec

Title: US-09-446-109a-6

Perfect score: 99

Sequence: 1 KYKHSVVKXXSEKPMPLXR 20

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	47.5	1676	1 C05_HUMAN	P01031 helio sapien
2	42	42.4	101	1 C05_HUMAN	P01093 helio sapien
3	42	42.4	602	1 VEL_CRPVK	P03112 cottontail
4	41	41.4	200	1 VIP_CHICK	P48143 gallus gall
5	41	41.4	200	1 VIP_MELGA	P45644 melalegris g
6	41	41.4	370	1 PCP1_AMPCA	P80484 amphidinium
7	41	41.4	1680	1 C05_MOUSE	P06684 mus musculus
8	40	40.4	257	1 MOVV_PPMVJ	P89658 pepper mild
9	40	40.4	257	1 MOVV_PPMVJ	P23097 pepper mild
10	39	39.4	613	1 VEL_PAPVD	P03117 deer papill
11	39	39.4	1451	1 DPOA_RAT	O89042 rattus norv
12	39	39.4	1465	1 DPOA_MOUSE	P33609 mus musculus
13	38.5	38.9	363	1 YK07_CAEEL	P34302 caenorhabdi
14	38	38.4	470	1 MTR_NEUCR	P38680 neurospora
15	38	38.4	506	1 DUPE_NPVAC	P18569 autographa
16	38	38.4	557	1 UVRK_THEMA	Q9WY43 thermotoga
17	37.5	37.9	1349	1 IKI3_YEAST	Q06706 saccharomyc
18	37	37.4	76	1 Y752_METJA	Q58162 methanococc
19	37	37.4	93	1 YARO_ERWHE	P42518 erwinia her
20	37	37.4	339	1 PYRC_HELPJ	Q92140 helicobacte
21	37	37.4	369	1 PCP2_AMPCA	P51873 amphidinium
22	37	37.4	429	1 THRC_BUCAI	P57289 buchnera ap
23	37	37.4	486	1 RBL1_RHOSH	P27997 rhodobacter
24	37	37.4	514	1 YHK8_YEAST	P38776 saccharomyc
25	37	37.4	522	1 CPL1_ONCMY	O92110 oncorhynch
26	37	37.4	629	1 GLDA_THEMA	Q9WY41 thermotoga
27	37	37.4	690	1 ORK3_SCHPO	Q63406 rattus norv
28	37	37.4	937	1 DBS_RAT	P14180 saccharomyc
29	37	37.4	963	1 CHS2_YEAST	Q09418 homo sapien
30	37	37.4	967	1 ATSL_HUMAN	P09499 caenorhabdi
31	37	37.4	983	1 YOG4_CAEEL	O15068 homo sapien
32	37	37.4	1108	1 DBS_HUMAN	Q64096 mus musculus
33	37	37.4	1149	1 DBS_MOUSE	

RESULT 1

ID	COS_HUMAN	STANDARD:	PRT:	1676 AA.
AC	P01031;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-DEC-1992 (Rel. 24, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	COMPLEMENT C5 PRECURSOR [CONTAINS: C5A ANAPHYLATOXIN].			
GN	C5.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91079575; PubMed=1984448;			
PA	Haviland D.L., Haviland J.C., Fleischer D.T., Hunt A., Wetsel R.A.;			
RT	"Complete cDNA sequence of human complement pro-C5. Evidence of			
RT	truncated transcripts derived from a single copy gene."			
RL	J. Immunol. 146:362-368(1991).			
RN	[2]			
RP	SEQUENCE OF 412-1676 FROM N.A.			
RX	MEDLINE=88209511; PubMed=3365401;			
RA	Wetsel R.A., Lemons R.S., Lebeau M.M., Barnum S.R., Noack D.,			
RA	Tack B.F.;			
RT	"Molecular analysis of human complement component C5: localization of			
RT	the structural gene to chromosome 9."			
RL	Biochemistry 27:1474-1482(1988).			
RN	[3]			
RP	SEQUENCE OF 412-902 FROM N.A.			
RX	MEDLINE=85130937; PubMed=2579066;			
RA	Lundwall A.B., Wetsel R.A., Kristensen T., Whitehead A.S.,			
RA	Woods D.E., Ogden R.C., Colten H.R., Tack B.F.;			
RT	"Isolation and sequence analysis of a cDNA clone encoding the fifth			
RT	complement component."			
RL	J. Biol. Chem. 260:2108-2112(1985).			
RN	[4]			
RP	SEQUENCE OF 678-751.			
RX	MEDLINE=79005687; PubMed=690134;			
RA	Fernandez H.N., Hugli T.E.;			
RT	"Primary structural analysis of the polypeptide portion of human C5a			
RT	anaphylatoxin. Polypeptide sequence determination and assignment of			
RT	the oligosaccharide attachment site in C5a."			
RL	J. Biol. Chem. 253:6955-6964(1978).			
RN	[5]			
RP	SEQUENCE OF 678-751 FROM N.A.			
RX	MEDLINE=91144547; PubMed=1996961;			
RA	Bohnsack J.F., Mollison K.W., Buko A.M., Ashworth J.C., Hill H.R.;			
RT	"Group B streptococci inactivate complement component C5a by enzymic			
RT	cleavage at the C-terminus."			
RL	Biochem. J. 273:635-640(1991).			
RN	[6]			
RP	STRUCTURE BY NMR OF C5a.			
RX	MEDLINE=88309754; PubMed=3408713;			
RA	Zuideweg E.R.P., Mollison K.W., Henkin J., Carter G.W.;			
RT	"Sequence-specific assignments in the 1H NMR spectrum of the human			

ALIGNMENTS


```
KW Thiol protease inhibitor.
SQ SEQUENCE 101 AA; 11186 MW; 789CD3E9611CBFD1 CRC64;

Query Match 42.4%; Score 42; DB 1; Length 101;
Best Local Similarity 47.4%; Pred. No. 1.7;
Matches 9; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

OY 2 YKHSVVKKXYSPKPLXR 20
   :|||  ||  |||  :
Db 85 WKHS--KKLLGFRPAPVDK 101

RESULT 3
VEI_CRPVK
ID VEI_CRPVK STANDARD; PRT: 602 AA.
AC P03112;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE REPLICATION PROTEIN EI.
GN Ei.
OS Cottontail rabbit (shope) papillomavirus (strain Kansas) (CRPV).
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=31553;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85166175; PubMed=2984661;
RA Giri I., Danos O., Yaniv M.;
RT 'Genomic structure of the cottontail rabbit (Shope) papillomavirus.';
RL Proc. Natl. Acad. Sci. U.S.A. 82:1580-1584(1985).
CC -I- FUNCTION: ATP-DEPENDENT DNA HELICASE REQUIRED FOR INITIATION OF
CC VIRAL DNA REPLICATION. IT FORMS A COMPLEX WITH THE VIRAL E2
CC PROTEIN. THE EI-E2 COMPLEX BINDS TO THE REPLICATION ORIGIN WHICH
CC CONTAINS BINDING SITES FOR BOTH PROTEINS.
CC -I- SUBCELLULAR LOCATION: NUCLEAR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K02708; -; NOT_ANNOTATED_CDS.
DR PIR: A03657; W1WLRB.
DR InterPro: IPR001177; Papillom_EI.
DR Pfam: PF00519; EI_1.
DR Pfam: PF00524; EI_N; 1.
KW Early protein; DNA replication; Helicase; ATP-binding; DNA-binding;
KW Nuclear protein.
FT NP_BIND 431 438 ATP (POTENTIAL).
SQ SEQUENCE 602 AA; 67943 MW; 239EF0F3F627D6A5 CRC64;

Query Match 42.4%; Score 42; DB 1; Length 602;
Best Local Similarity 43.8%; Pred. No. 11;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 YKHSVVKKXYSPKMP 17
   ||  :|  :|  |||
Db 275 YKRAMVSGVSYGMP 290

RESULT 4
VIP_CHICK
ID VIP_CHICK STANDARD; PRT: 200 AA.
AC P48143; P01285;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
```

```
DE VASOACTIVE INTESTINAL PEPTIDE PRECURSOR (VIP).
GN VIP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA McFarlin D.R., Lehn D.A., Moran S.M., Macdonald M.J., Epstein M.L.;
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-84 AND 120-200 FROM N.A.
RC TISSUE=Hypothalamus;
RA Talbot R.T., Dunn I.C., Wilson P.W., Sang H.M., Sharp P.J.;
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 129-156.
RX MEDLINE=76210823; PubMed=1227973;
RA Nilsson A.;
RT "Structure of the vasoactive intestinal octacosapeptide from chicken
RT intestine. The amino acid sequence.";
RL FEBS Lett. 60:322-326(1975).
RN [4]
RP SYNTHESIS OF VIP, AND SEQUENCE OF 139-156.
RA Bodanszky M., Lin C.Y., Yiotakis A.E., Mutt V., Said S.I.;
RT "Vasoactive intestinal peptide (VIP) from chicken. Synthesis and
RT properties of the C-terminal hendecapeptide.";
RL Bioorg. Chem. 5:339-350(1976).
CC -I- FUNCTION: VIP CAUSES VASODILATION, LOWERS ARTERIAL BLOOD
CC PRESSURE, STIMULATES MYOCARDIAL CONTRACTILITY, INCREASES
CC GLYCOGENOLYSIS AND RELAXES THE SMOOTH MUSCLE OF TRACHEA, STOMACH
CC AND GALL BLADDER.
CC -I- FUNCTION: PHI ALSO CAUSES VASODILATION.
CC -I- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09350; AAA87896.1; -
DR EMBL: X80906; CAA56867.1; -
DR PIR: A01551; VRCH.
DR InterPro: IPR000532; Glucagon.
DR Pfam: PF00123; hormone2; 2.
DR PRINTS: PR00275; GLUCAGON.
DR SMART: SM00070; GLUCA; 2.
DR PROSITE: PS00260; GLUCAGON; 1.
KW Glucagon family; Cleavage on pair of basic residues; Signal;
KW Amidation; Hormone; Alternative splicing.
FT SIGNAL 1 25 POTENTIAL.
FT PEPTIDE 89 115 INTESTINAL PEPTIDE PHI-27.
FT PEPTIDE 129 156 VASOACTIVE INTESTINAL PEPTIDE.
FT MOD_RES 156 156 AMIDATION (G-157 PROVIDE AMIDE GROUP).
FT VARSPPLIC 85 119 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 200 AA; 22539 MW; 902A88F998CAB402 CRC64;

Query Match 41.4%; Score 41; DB 1; Length 200;
Best Local Similarity 35.0%; Pred. No. 5.2;
Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 1 KYKHSVVKKXYSPKPLXR 20
   :|  ||::|  |  :|  |
Db 109 RYLHSLIRKRVSSQDSPVKR 128

RESULT 5
```

VIP_MELGA STANDARD; PRT: 200 AA.

ID VIP_MELGA

AC P45644;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE VASOACTIVE INTESTINAL PEPTIDE PRECURSOR (VIP).

GN VIP.

OS Meleagris gallopavo (Common turkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Meleagrididae; Meleagris.

OX NCBI_TaxID=9103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-NICHOLAS WHITE; TISSUE-HYPOTHALAMUS;

RX MEDLINE=95269705; PubMed=7750483;

RA You S., Salsby J.L., Farris J.A., Foster D.N., el Halawani M.E.;

RT "Tissue-specific alternative splicing of turkey preprovasoactive intestinal peptide messenger ribonucleic acid, its regulation, and correlation with prolactin secretion.";

RT Endocrinology 136:2602-2610(1995)

RL

CC -!- FUNCTION: VIP CAUSES VASODILATION, LOWERS ARTERIAL BLOOD PRESSURE, STIMULATES MYOCARDIAL CONTRACTILITY, INCREASES GLYCOGENOLYSIS AND RELAXES THE SMOOTH MUSCLE OF TRACHEA, STOMACH AND GALL BLADDER.

CC -!- FUNCTION: PHI ALSO CAUSES VASODILATION.

CC -!- SIMILARITY: BELONGS TO THE GLUCAGON FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: L36641; AAA92866.1;

DR InterPro: IPR000532; Glucagon.

DR Pfam: PF00123; hormone2; 2.

DR PRINTS: PR00275; GLUCAGON.

DR SMART: SM00070; GLUCA; 2.

DR PROSITE: PS00260; GLUCAGON; 1.

DR Glucagon family; Cleavage on pair of basic residues; Signal;

KW Amidation; Hormone.

FT SIGNAL 1 25 POTENTIAL.

FT PEPTIDE 89 115 INTESTINAL PEPTIDE PHI-27.

FT PEPTIDE 129 156 VASOACTIVE INTESTINAL PEPTIDE.

FT MOD_RES 156 156 AMIDATION (G-157 PROVIDE AMIDE GROUP).

SQ SEQUENCE 200 AA; 22600 MW; B35933495273F06D CRC64;

Query Match 41.4%; Score 41; DB 1; Length 200;
Best Local Similarity 35.0%; Pred. No. 5.2;
Matches 7; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 1 KYKHSVKKXYSFKPMPLXR 20
:|||||:|:|:|

DB 109 RYLHSLIRKRVSSQDSQPVKR 128

RESULT 6

PCP1_AMPCA STANDARD; PRT: 370 AA.

ID PCP1_AMPCA

AC P80484; P51872;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PERIDININ-CHLOROPHYLL A PROTEIN 1 PRECURSOR (PCP).

OS Amphidinium carterae (Dinoflagellate).

OC Eukaryota; Alveolata; Dinophyceae; Gymnodiniales; Gymnodiniaceae;

OC Amphidinium.

OX NCBI_TaxID=2961;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN-CS-21;

RX MEDLINE=96413810; PubMed=8816945;

RA Sharples F.P., Wrench P.M., Ou K., Hiller R.G.;

RT "Two distinct forms of the peridinin-chlorophyll a-protein from Amphidinium carterae.";

RL Biochim. Biophys. Acta 1276:117-123(1996).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).

RX MEDLINE=96256477; PubMed=8650577;

RA Hofmann E., Wrench P.M., Sharples F.P., Hiller R.G., Welte W., Diederichs K.;

RT "Structural basis of light harvesting by carotenoids: peridinin-chlorophyll-protein from Amphidinium carterae.";

RL Science 272:1788-1791(1996).

CC -!- FUNCTION: WATER-SOLUBLE ANTENNA FOR CAPTURE OF SOLAR ENERGY IN THE BLUE-GREEN RANGE. PERIDININ IS AN ASYMMETRIC CAROTENOID HAVING A MAXIMUM ABSORPTION AT APPROXIMATELY 480 NM.

CC -!- SUBUNIT: HOMOTRIMER.

CC -!- DOMAIN: THE MATURE PROTEIN IS COMPOSED OF 2 ALMOST IDENTICAL REPEAT UNITS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL: 250792; CAA90653.1;

DR EMBL: X94549; CAA64242.1;

DR PDB: 1PPR; 20-AUG-97.

DR InterPro: IPR003376; PCP.

DR Pfam: PF02429; PCP; 2.

KW Light-harvesting polypeptide; Transit peptide; Chloroplast; Multigene family; Repeat; 3D-structure.

FT TRANSIT 1 57 CHLOROPLAST.

FT CHAIN 58 370 PERIDININ-CHLOROPHYLL A PROTEIN 1.

FT SITE 123 123 CHLOROPHYLL A-BINDING.

FT SITE 286 286 CHLOROPHYLL A-BINDING.

FT DOMAIN 37 48 POLY-ALA.

FT REPEAT 58 220 1.

FT REPEAT 221 370 2.

FT CONFLICT 73 73 A -> H (IN REF. 1; CAA64242).

FT CONFLICT 98 98 L -> T (IN REF. 1; CAA64242).

FT CONFLICT 130 130 V -> I (IN REF. 1; AA SEQUENCE).

FT CONFLICT 243 243 W -> E (IN REF. 1; AA SEQUENCE).

SQ SEQUENCE 370 AA; 38159 MW; 7F113EB0D8F372E CRC64;

Query Match 41.4%; Score 41; DB 1; Length 370;
Best Local Similarity 57.1%; Pred. No. 9.7;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 SVVKXYSFKPMPL 18
|||||:|:|:|

DB 20 SVVQKSHGEVPSPL 33

RESULT 7

C05_MOUSE STANDARD; PRT: 1680 AA.

ID C05_MOUSE

AC P06684;

DT 01-JAN-1988 (Rel. 06, Created)

DT 01-NOV-1990 (Rel. 16, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE COMPLEMENT C5 PRECURSOR (HEMOLYTIC COMPLEMENT) [CONTAINS: C5A DE ANAPHYLATOXIN].

GN C5 OR HC.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=90153853; PubMed=2303408;
RA Wetzel R.A., Fleischer D.T., Haviland D.L.;
RT "Deficiency of the murine fifth complement component (C5). A 2-base
RT pair gene deletion in a 5'-exon";
RL J. Biol. Chem. 265:2435-2440(1990).
RN [2]
RN SEQUENCE OF 41-1680 FROM N.A.
RX MEDLINE=87185363; PubMed=2436653;
RA Wetzel R.A., Ogata R.T., Tack B.F.;
RT "Primary structure of the fifth component of murine complement.";
RL Biochemistry 26:737-743(1987).
CC -!- FUNCTION: ACTIVATION OF C5 BY A C5 CONVERTASE INITIATES THE
CC SPONTANEOUS ASSEMBLY OF THE LATE COMPLEMENT COMPONENTS, C5-C9,
CC INTO THE MEMBRANE ATTACK COMPLEX. C5B HAS A TRANSIENT BINDING SITE
CC FOR C6. THE C5B-C6 COMPLEX IS THE FOUNDATION UPON WHICH THE LYtic
CC COMPLEX IS ASSEMBLED.
CC -!- FUNCTION: DERIVED FROM PROTEOLYTIC DEGRADATION OF COMPLEMENT C5,
CC C5 ANAPHYLATOXIN IS A MEDIATOR OF LOCAL INFLAMMATORY PROCESS. IT
CC INDUCES THE CONTRACTION OF SMOOTH MUSCLE, INCREASES VASCULAR
CC PERMEABILITY AND CAUSES HISTAMINE RELEASE FROM MAST CELLS AND
CC BASOPHILIC LEUKOCYTES. C5A ALSO STIMULATES THE LOCOMOTION OF
CC POLYMORPHONUCLEAR LEUKOCYTES (CHEMOKINESIS) AND DIRECT THEIR
CC MIGRATION TOWARD SITES OF INFLAMMATION (CHEMOTAXIS).
CC -!- SUBUNIT: C5 PRECURSOR IS FIRST PROCESSED BY THE REMOVAL OF 4 BASIC
CC RESIDUES, FORMING TWO CHAINS, BETA & ALPHA, LINKED BY A DISULFIDE
CC BOND. C5 CONVERTASE ACTIVATES C5 BY CLEAVING THE ALPHA CHAIN,
CC RELEASING C5A ANAPHYLATOXIN & GENERATING C5B (BETA CHAIN + ALPHA
CC CHAIN).
CC -!- SIMILARITY: TO C3, C4 AND ALPHA-2-MACROGLOBULIN.
CC -!- SIMILARITY: CONTAINS 1 ANAPHYLATOXIN-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DR ENBL: M35525; AAA37349.1; -;
DR ENBL: M35526; AAA37348.1; -;
DR PIR: A27538; A27538.
DR PIR: A35530; A35530.
DR HSSP: P01031; 1KJS.
DR MGD: MGI:96011; HC.
DR InterPro: IPR002890; A2M_N.
DR InterPro: IPR001599; Alpha_2_macroloblin.
DR InterPro: IPR000020; Anaphylatoxin.
DR InterPro: IPR001840; Anaphylatoxin.
DR InterPro: IPR001134; Netrin_C.
DR Pfam: PF00207; A2M; 1.
DR Pfam: PF01835; A2M_N; 1.
DR Pfam: PF01821; ANATO; 1.
DR Pfam: PF01759; NTR; 1.
DR SMART: PR00004; ANAPHYLATOXN.
DR SMART: SH00104; ANATO; 1.
DR PROSITE: PS00477; ALPHA_2_MACROGLOBULIN; FALSE_NEG.
DR PROSITE: PS01177; ANAPHYLATOXIN_1; 1.
DR PROSITE: PS01178; ANAPHYLATOXIN_2; 1.
KW Complement pathway: Complement alternate pathway: Glycoprotein;
KW Plasma: Membrane attack complex: Cytolysis; Inflammatory response;
KW Signal.
FT SIGNAL 1 18 COMPLEMENT C5.
FT CHAIN 19 1680 COMPLEMENT C5 BETA CHAIN.
FT CHAIN 19 674
FT PROPEP 675 678
FT CHAIN 679 1680 COMPLEMENT C5 ALPHA CHAIN.
FT PEPTIDE 679 755 C5A ANAPHYLATOXIN.
FT CHAIN 756 1680 C5B (ALPHA').
FT DOMAIN 702 736 ANAPHYLATOXIN-LIKE.

FT DISULFID 702 728 BY SIMILARITY.
FT DISULFID 703 735 BY SIMILARITY.
FT CARBOHYD 715 736 BY SIMILARITY.
FT CARBOHYD 427 427 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 915 915 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1119 1119 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1633 1633 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 216 216 Y -> L (IN DEFECTIVE VARIANT C5D).
FT VARIANT 217 1680 MISSING (IN DEFECTIVE VARIANT C5D).
SQ SEQUENCE 1680 AA; 188877 MW; 81E5A16FAC7D95C CRC64;

Query Match 41.4%; Score 41; DB 1; Length 1680;
Best Local Similarity 88.9%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 KYKHSVVKK 9
DB 693 KYKHSVVKK 701
IIIIII II

RESULT 8
MOV_PPMVJ STANDARD; PRT; 257 AA.
ID MOV_PPMVJ
AC P89658;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MOVEMENT PROTEIN (CELL-TO-CELL TRANSPORT PROTEIN) (30 KDA PROTEIN).
GN MP.
OS Pepper mild mottle virus (strain Japan) (PMV-J).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138663;
RN [1]
RN SEQUENCE FROM N.A.
KA Kirita M., Akutsu K., Watanabe Y., Tsuda S.;
RT "Nucleotide sequence of the Japanese isolate of pepper [Capsicum
RT annum] mild mottle tobamovirus (IMV-P) RNA";
RL Ann. Phytopathol. Soc. Jpn. 63:373-376(1997).
CC -!- FUNCTION: INVOLVED IN TRANSPORT OF THE VIRUS FROM THE INITIALLY
CC INFECTED CELLS TO ADJACENT CELLS, POSSIBLY BY MODIFYING THE
CC FUNCTION OF THE PLASMODESMATA. ALSO INFLUENCES LOCAL LESION
CC DEVELOPMENT. BINDS TO SINGLE-STRANDED NUCLEIC ACID.
CC -!- SIMILARITY: BELONGS TO THE TOBAMOVIRUSES MOVEMENT PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@lsb-sib.ch).
CC -----
DR ENBL: AB000709; BAA19168.1; -;
DR InterPro: IPR001022; Tobamo_MP.
DR Pfam: PF01107; Tobamo_MP; 1.
DR PRINTS: PR00964; MOVEMENT.
KW DNA-binding; Transport.
SQ SEQUENCE 257 AA; 28345 MW; 9F375F2E691FB0A7 CRC64;

Query Match 40.4%; Score 40; DB 1; Length 257;
Best Local Similarity 43.8%; Pred. No. 10;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 YKHSVVKKXYSFKPMP 17
DB 108 YRTSAKKRFAFKLIP 123
I: I II::II:I

RESULT 9
MOV_PPMVJ STANDARD; PRT; 257 AA.
ID MOV_PPMVJ

DR SMART: SM00486; POLBC; 1.
 DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
 KW Transferase: DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding; Nuclear protein.
 FT DNA_BIND 657 722 POTENTIAL.
 FT DNA_BIND 1252 1383 POTENTIAL.
 FT NON_TER 1451 1451
 SQ SEQUENCE 1451 AA; 165305 MW; F0E1B16F8BBD5CD2 CRC64;

Query Match: 39.4%; Score 39; DB 1; Length 1451;
 Best Local Similarity 42.9%; Pred. No. 89;
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 KYKHSVVKXYSK 14
 I: I: I: I: I: I:
 Db 436 KPKSKIVKKNYAF 449

RESULT 12
 DPOA_MOUSE
 ID DPOA_MOUSE STANDARD: PRT: 1465 AA.
 AC P33609;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA POLYMERASE ALPHA CATALYTIC SUBUNIT (EC 2.7.7.7).
 GN POLA OR POLAL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 454-472 AND 1438-1455.
 RX MEDLINE=93216788; PubMed=8463324;
 RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M.,
 RA Hanaoka F.;
 RT "Molecular cloning of the cDNAs for the four subunits of mouse DNA
 RT polymerase alpha-primase complex and their gene expression during
 RT cell proliferation and the cell cycle.";
 RL J. Biol. Chem. 268:8111-8122(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94298818; PubMed=8026492;
 RA Stadlbauer F., Brueckner A., Rehfuess C., Eckerskorn C.,
 RA Lottspeich F., Foerster V., Tseng B.Y., Nasheuer H.P.;
 RT "DNA replication in vitro by recombinant
 RT DNA-polymerase-alpha-primase.";
 RL Eur. J. Biochem. 222:781-793(1994).
 RN [3]
 RP VARIANT PHE-1180.
 RX MEDLINE=94171797; PubMed=8125989;
 RA Izumi M., Miyazawa H., Harakawa S., Yatagai F., Hanaoka F.;
 RT "Identification of a point mutation in the cDNA of the catalytic
 RT subunit of DNA polymerase alpha from a temperature-sensitive mouse
 RT FM3A cell line.";
 RL J. Biol. Chem. 269:7639-7644(1994).
 CC -1- FUNCTION: POLYMERASE ALPHA IN A COMPLEX WITH DNA PRIMASE IS A
 CC REPLICATIVE POLYMERASE.
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
 CC N PYROPHOSPHATE + DNA(N).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
 CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
 CC DIFFERENT REACTIONS OF DNA SYNTHESIS.
 CC -1- MISCELLANEOUS: CONSERVED REGIONS II, IV, III AND I ARE THOUGHT TO
 CC BE INVOLVED IN SUBSTRATE RECOGNITION, BINDING OR PP(I) HYDROLYSIS.
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L16559; AAA27932.1; -;
 CC WormPep; C06E1.7; CE00062.

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D13543; BAA40003.1; -;
 DR EMBL: D17384; BAA04202.1; -;
 DR PIR: S45628; S45628.
 DR MGI: 99660; Polal.
 DR InterPro: IPR002064; DNA_pol_B.
 DR Pfam: PF00136; DNA_pol_B; 1.
 DR PRINTS: PR00106; DNAPOLB.
 DR SMART: SM00486; POLBC; 1.
 DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
 KW Transferase: DNA-directed DNA polymerase; DNA replication;
 KW DNA-binding; Nuclear protein.
 FT DNA_BIND 654 719 POTENTIAL.
 FT DNA_BIND 1249 1380 POTENTIAL.
 FT VARIANT 1180 1180
 FT S -> F (IN TEMPERATURE-SENSITIVE FT20
 FT CELL LINE: DEFECTIVE ACTIVITY).
 SQ SEQUENCE 1465 AA; 167339 MW; 46D1A9818A7944A3 CRC64;

Query Match: 39.4%; Score 39; DB 1; Length 1465;
 Best Local Similarity 42.9%; Pred. No. 90;
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 KYKHSVVKXYSK 14
 I: I: I: I: I: I:
 Db 433 KPKSKIVKKNYAF 446

RESULT 13
 YKQ7_CAEEL
 ID YKQ7_CAEEL STANDARD: PRT: 363 AA.
 AC P34302;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE HYPOPHETICAL 41.7 KDA PROTEIN C06E1.7 IN CHROMOSOME III.
 GN C06E1.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
 RA Sims M., Smaldon N., Smith A., Smith K., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L16559; AAA27932.1; -;
 CC WormPep; C06E1.7; CE00062.

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M22619; AAA69845.1; -;
 DR EMBL; L09723; AAA46712.1; -;
 DR EMBL; M96361; AAA66785.1; -;
 DR EMBL; L22858; AAA66645.1; -;
 DR PIR; A34114; XUNVAC.
 DR InterPro: IPR002213; UDPGT.
 DR Pfam; PF00201; UDPGT; 1.
 DR PROSITE; PS00375; UDPGT; 1.
 KW Transferase: Glycosyltransferase; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 506 ECDYSTEROID UDP-GLUCOSYLTRANSFERASE.
 SQ SEQUENCE 506 AA; 57033 MW; C83FA34D2453A606 CRC64;

Query Match 38.4%; Score 38; DB 1; Length 506;
 Best Local Similarity 35.7%; Pred. No. 45;
 Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 1 KYKHSVVKXXYSPK 14
 :|: :|: |
 DB 465 RYRHDISRQLYSLK 478

Search completed: February 27, 2002, 11:42:44
 Job time: 543 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:06 ; Search time 281.76 Seconds
(without alignments)
10.383 Million cell updates/sec

Title: US-09-446-109A-6
 Perfect score: -99
 Sequence: 1 KYKHSVVVKXXYSFKPMPLXR 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database :
SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	51.5	241	12	Q84674 paramescium
2	50	50.5	610	5	Q9n505 caenorhabdi
3	48	48.5	2219	5	Q23388 caenorhabdi
4	46.5	47.0	1279	2	Q26046 helicobacte
5	45	45.5	1170	5	Q22624 caenorhabdi
6	44	44.4	1313	2	Q9PQ16 ureaplasma
7	43	43.4	346	1	Q9YDW6 aeropyrum p
8	43	43.4	1579	5	Q9GR77 leishmania
9	42.5	42.9	112	12	Q9YR54 aleutian mi
10	42	42.4	479	2	Q9PJ47 campylobact
11	42	42.4	510	2	Q9RA39 trichodesmi
12	42	42.4	593	5	Q9VJQ2 drosophila
13	42	42.4	602	12	Q9IES2 cottontail
14	42	42.4	602	12	Q9ICK9 cottontail
15	42	42.4	872	11	Q55000 rattus norv
16	42	42.4	940	4	Q00405 homo sapien
17	41.5	41.9	2328	5	Q9vw47 drosophila
18	41.5	41.9	2531	5	Q9GPH4 drosophila
19	41	41.4	3730	10	Q9fpw2 amphidinium

Q949414 arabidopsis
Q91949 caenorhabdi
Q919479 drosophila
Q921997 streptococ
Q67015 auxilif ase
Q9413 leishmania
Q949418 yaba-like d
Q949418 pseudomonas
Q949418 bacillus ha
Q949418 candida alb
Q949418 candida alb
Q949418 caenorhabdi
Q949418 mus musculu
Q949418 saccharomyc
Q949418 bacterioph
Q949418 mastigocid
Q949418 helicoverp
Q949418 arabidopsi
Q949418 oryza sativ
Q949418 pseudomonas
Q949418 vibrio chol
Q949418 vibrio chol
Q949418 drosophila
Q949418 drosophila
Q949418 caenorhabdi
Q949418 thermotoqa

ALIGNMENTS

```

RESULT 1
ID Q84674 PRELIMINARY; PRT; 241 AA.
AC Q84674;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE GENOME, PARTIAL SEQUENCE.
GN A360R.
OS Paramoecium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae.
OX NCBI_TaxID=10506;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=95133167; PubMed=7831789;
RT Lu Z., Li Y., Zhang Y., Kutish G.F., Rock D.L., Van Et
RA "Analysis of 45 kb of DNA located at the left end of t
RL virus PBCV-1 genome.";
RL Virology 206:339-352(1995).
DR EMBL: U42580; AAC96728.1;
DS SEQUENCE 241 AA; 26311 MW; A47481C943BE5015 CRC64;

Query Match 51.5%; Score 51; DB 12; Length
Best Local Similarity 52.6%;
Matches 10; Conservative 2; Mismatches 7; Index

QY 2 YKHSVVKXXYSFKPMLXR 20
:||||| ||:|
Db 21 FKRSVVKKPLGIKPLPARR 39

RESULT 2
ID Q9N505 PRELIMINARY; PRT; 610 AA.
AC Q9N505;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL PROTEIN Y40C7B.5.
GN Y40C7B.5.
OS Caenorhabditis elegans.

```



```
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Jones K., Tin-Wollam A., Yoakum M.;
RT "The sequence of C. elegans cosmid Y40C7B."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AC006757; AAF60545.1;
SQ SEQUENCE 610 AA; 70271 MW; D7C20E3F77457816 CRC64;

Query Match 50.5%; Score 50; DB 5; Length 610;
Best Local Similarity 50.0%; Pred. No. 2.8;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 KHSVVKKXYSFKPMPPLXR 20
DB 67 KHKIVETLNNFKPSPLOLR 84

RESULT 3
Q23388 PRELIMINARY; PRT: 2219 AA.
AC Q23388;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE 2K1067.2 PROTEIN.
DE 2K1067.2
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Thomas K.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBSJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Wilson R.J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smaildon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; 270038; CAA93884.1;
DR InterPro; IPR000169; Tholprot_act_site.
DR InterPro; IPR000967; Znf_NFX1.
DR InterPro; IPR003015; HLH_MYC.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.

DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
DR SMART; SMO0436; Znf_NFX_4.
SQ SEQUENCE 2219 AA; 253649 MW; F0F5A0F37D04ED0C CRC64;

Query Match 48.5%; Score 48; DB 5; Length 2219;
Best Local Similarity 47.6%; Pred. No. 23;
Matches 10; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 2 YKHSV--VKKXYSFKPMPPLXR 20
DB 666 YKHVLENLKRISSEKPLPFER 686

RESULT 4
Q26046 PRELIMINARY; PRT: 1279 AA.
AC Q26046;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE TYPE IIS RESTRICTION ENZYME R AND M PROTEIN (ECO57IR).
DE HP1517.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
DR EMBL; AE000649; AAO08556.1;
DR TIGR; HP1517;
DR InterPro; IPR002296; N12N6_mtfrase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1279 AA; 149714 MW; 5838BF4E242A7A31 CRC64;

Query Match 47.0%; Score 46.5; DB 2; Length 1279;
Best Local Similarity 45.5%; Pred. No. 23;
Matches 10; Conservative 5; Mismatches 4; Indels 3; Gaps 1;

QY 2 YKHSVVKKXYSF---KPMPLXR 20
DB 369 YKNSVLKHKHKDYQKEKPLPLK 390

RESULT 5
Q22624 PRELIMINARY; PRT: 1170 AA.
AC Q22624;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE T21B10.3 PROTEIN.
DE T21B10.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
```

OC	NCBI_TaxID=56636;
OX	[1]
RN	SEQUENCE FROM N.A.
KP	STRAIN=K1;
RC	MEDLINE=99310319; PubMed=10382966;
RX	Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Hoikawa Y.,
RA	Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA	Hosoyana A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA	Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA	Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA	Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT	"Complete genome sequence of an aerobic hyper-thermophilic
RT	Crenarchaeon, Aeropyrum pernix K1.";
RL	DNA Res. 6:83-101(1999).
RL	EMBL; AP000060; BAA79781.1; -
KW	Hypothetical protein; Complete proteome.
DW	SEQUENCE 346 AA; 36772 MW; 3DAB2DBD936EC3DF CRC64;
SQ	
Query Match	43.4%; Score 43; DB 1; Length 346;
Best Local Similarity	66.7%; Pred. No. 24;
Matches	8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY	6 VVKXXYSFKPMPLR 17
QB	12 VVKGTYSLKPLP 23
RESULT	8
U9GRT7	
ID	PRELIMINARY; PRT; 1579 AA.
AC	O9GRT7;
DT	01-MAR-2001 (TrEMBLrel. 16, Created)
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
LT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	PUTATIVE MITOGEN-ACTIVATED PROTEIN KINASE 8.
GN	MPK6.
OS	Leishmania mexicana.
OC	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX	NCBI_TaxID=5665;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=MNYC/BZ/62/M379;
RT	Wiese M., Wang Q., Goercke I.;
RT	"Identification of mitogen-activated protein (MAP) kinases from
RT	Leishmania mexicana."
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AJ293286; CAC07962.1; -
DR	InterPro; IPR000719; Euk_pkinase.
DR	InterPro; IPR002290; Ser_thr_kin_actsite.
DR	InterPro; IPR001245; Tyr_kin.
DR	pfam; pf00069; pkinase; 2.
DR	SMART; SM00220; S_TKC; 1.
DR	SMART; SM00219; TyrKC; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW	ATP-binding; Kinase; Transferase.
SQ	SEQUENCE 1579 AA; 165702 MW; DC2630F2634D8736 CRC64;
Query Match	43.4%; Score 43; DB 5; Length 1579;
Best Local Similarity	50.0%; Pred. No. 1.e+02;
Matches	8; Conservative 3; Mismatches 5; Indels 0; Caps 0;
QY	5 SVVKXXYSFKPMPLXR 20
	::: :
QB	774 SLKXYSFKSIPTSR 789
RESULT	9
G9YR54	
Q9YR54	PRELIMINARY; PRT; 112 AA.

```
AC Q9YR54;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2001 (TrEMBLrel. 16, Last annotation update)
DE NONSTRUCTURAL PROTEIN 1 (FRAGMENT).
GN NS1.
OS Aleutian mink disease virus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Parvovirus.
OX NCBI_TaxID=28314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El;
RA Olofsson A., Mittelholzer C., Treiberg Berndtsson L., Lind L.,
RA Mejerland T., Belak S.;
RT "Unusual high genetic diversity of Aleutian mink disease virus.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF107634; AAD04600.1; -
KW Nonstructural protein.
FT NON_TER 1
FT 112 112
SQ SEQUENCE 112 AA; 13052 MW; 0DFBAC185243FF6F CRC64;

Query Match 42.9%; Score 42.5; DB 12; Length 112;
Best Local Similarity 50.0%; Pred. No. 9.3;
Matches 9; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 KYKHSVVKXXYSFKPMP L 8
   || | | | | | | | |
DB 67 KYHRRQTKODYN-KPVP L 83

RESULT 10
Q9PJ47 PRELIMINARY: PRT: 479 AA.
AC Q9PJ47;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE IRON-SULFUR PROTEIN.
GN CJO074C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrall B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139074; CAB72560.1; -
DR InterPro: IPR001450; 4FE4S.ferredxin.
DR Pfam: PF00037; fer4; 2.
DR PROSITE: PS00198; 4FE4S.FERREDOXIN; 2.
KW Complete proteome.
SQ SEQUENCE 479 AA; 54220 MW; 3BB1885D5C9F88F8 CRC64;

Query Match 42.4%; Score 42; DB 2; Length 479;
Best Local Similarity 58.3%; Pred. No. 50;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 6 VVKXXYSFKPMP L 7
   | | | | | | | |

```

```
Lb 449 VIKKWYAFKELP 460

RESULT 11
Q9RA39 PRELIMINARY: PRT: 510 AA.
ID Q9RA39;
AC Q9RA39;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
LT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FERREDOXIN NITRITE REDUCTASE.
GN NIRA.
OS Trichodesmium sp. WH9601.
OC Bacteria; Cyanobacteria; Oscillatoriales; Trichodesmium.
OX NCBI_TaxID=104582;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WH9601;
RA Li H., Post A.F.;
RT "The nitrate assimilation genes of the diazotrophic, bloom-forming
RT cyanobacterium Trichodesmium strain WH9601.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF178846; AAF00916.1; -
DR InterPro: IPR000660; Nir_Sir.
DR Pfam: PF01077; NIR_SIR; 1.
DR PRINTS: PR00397; SIROHAEM.
DR PROSITE: PS00365; NIR_SIR; 1.
SQ SEQUENCE 510 AA; 57584 MW; F3ED1B8E0BC8619F CRC64;

Query Match 42.4%; Score 42; DB 2; Length 510;
Best Local Similarity 36.8%; Pred. No. 53;
Matches 7; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

CY 2 YKHSVVKXXYSFKPMP LXR 20
   : : | : | : | | | |
DB 373 PRSEPLKKFSPEQP LTR 391

RESULT 12
Q9VJ02 PRELIMINARY: PRT: 593 AA.
ID Q9VJ02;
AC Q9VJ02;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
IT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BG:DS04929.3 OR CG15269.
GN BG:DS04929.3 OR CG15269.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bröttier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
```

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.W., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL: AE003646; AAF53440.1; -.
DR HSP: P08047; 1SP2.
DR FlyBase: FBgn028878; BC:DS04929.3.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; Zf-C2H2; 8.
DR PRINTS: PR00048; ZINC_FINGER.
DR SMART: SM00355; Znf_C2H2; 8.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
KW DNA-binding; Metal-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 593 AA: 64409 MW: 347CE5F8138B2EE2 CRC64;

Query Match 42.4%; Score 42; DB 5; Length 593;
Best Local Similarity 31.2%; Pred. No. 62;
Matches 5; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 YKHSVVKXYSFKPMP 17
:||||:| | |
DB 3 FXHMLQRYDLSPSP 18

RESULT 13
Q9IES2 PRELIMINARY; PRT; 602 AA.
AC Q9IES2
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE COTTONTAIL RABBIT PAPILLOMAVIRUS SUBTYPE B COMPLETE GENOME.
GN El.
OS Cottontail rabbit papillomavirus.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SUBTYPE B;
RA Salmon J., Orth G., Breitbart F.;
RT "Complete nucleotide sequences of the CRPVb strain."
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ243287; CAB96116.1; -.
DR InterPro: IPR001177; Papillom_El.
DR Pfam: PF00519; El; 1.
DR Pfam: PF00524; El_N; 1.
SQ SEQUENCE 602 AA: 68015 MW: 8458F4C93DAlF64B CRC64;

Query Match 42.4%; Score 42; DB 12; Length 602;
Best Local Similarity 43.8%; Pred. No. 63;

Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY 2 YKHSVVKXYSFKPMP 17
||:| | | |
DB 275 YKRAMVSGVFSYGPMP 290

RESULT 14

Q9ICK9 PRELIMINARY; PRT; 602 AA.
AC Q9ICK9
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE A4 COMPLETE GENOME.
GN El.
OS Cottontail rabbit papillomavirus.
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A4;
RA Salmon J.;
RT "CRPVa4 complete nucleotide sequence";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ404003; CAB96164.1; -.
DR InterPro: IPR001177; Papillom_El.
DR Pfam: PF00519; El; 1.
DR Pfam: PF00524; El_N; 1.
SQ SEQUENCE 602 AA: 67904 MW: A9B3382E4533CACB CRC64;

Query Match 42.4%; Score 42; DB 12; Length 602;
Best Local Similarity 43.8%; Pred. No. 63;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 2 YKHSVVKXYSFKPMP 17
||:| | | |
DB 275 YKRAMVSGVFSYGPMP 290

RESULT 15

O55000 PRELIMINARY; PRT; 872 AA.
AC O55000;
DT 01-JUN-1998 (TremBLrel. 06, Created)
DT 01-JUN-1998 (TremBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE HYPOTHETICAL 92.8 KDA PROTEIN.
GN PNUTS.
OS Rattus norvegicus (rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RA Allen P.B., Ouimet C.C., Greengard P.;
RL J. Biol. Chem. 0:0-0(1998).
DR EMBL: AF040954; AAB96775.1; -.
DR InterPro: IPR000571; Zf-CCCH.
DR InterPro: IPR003617; TFS2_N.
DR SMART: SM00509; TFS2N; 1.
DR SMART: SM00356; Znf_C3H1; 1.
KW Hypothetical protein.
SQ SEQUENCE 872 AA: 92828 MW: 23CC61B4B296E948 CRC64;

Query Match 42.4%; Score 42; DB 11; Length 872;
Best Local Similarity 53.3%; Pred. No. 91;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 VVKKXYSFKPMLXR 20
II : II:II I
Db 231 VVSDKYNLRPIPLKR 245

Search completed: February 27, 2002, 11:50:07
Job time: 986 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:03 ; Search time 132.19 Seconds
(without alignments)
3.405 Million cell updates/sec

Title: US-09-446-109A-6
Perfect score: 99
Sequence: 1 KYKHSVKKXYSFKPMLXR 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	47	47.5	74	1	US-08-463-224-3	Sequence 3, Appli
2	47	47.5	74	2	US-08-463-377-3	Sequence 3, Appli
3	47	47.5	74	4	US-09-246-5008-10	Sequence 10, Appl
4	47	47.5	77	4	US-09-246-5008-14	Sequence 14, Appl
5	42	42.4	10	1	US-08-299-285-7	Sequence 7, Appli
6	42	42.4	10	1	US-08-299-285-10	Sequence 10, Appl
7	42	42.4	10	1	US-08-299-285-24	Sequence 24, Appl
8	42	42.4	10	2	US-08-985-126-7	Sequence 7, Appli
9	42	42.4	10	2	US-08-985-126-10	Sequence 10, Appl
10	42	42.4	10	2	US-08-985-126-24	Sequence 24, Appl
11	42	42.4	10	5	PCT-US95-11126-7	Sequence 7, Appli
12	42	42.4	10	5	PCT-US95-11126-10	Sequence 10, Appl
13	42	42.4	10	5	PCT-US95-11126-24	Sequence 24, Appl
14	41	41.4	10	1	US-08-299-285-5	Sequence 5, Appli
15	41	41.4	10	1	US-08-299-285-13	Sequence 13, Appl
16	41	41.4	10	2	US-08-985-126-5	Sequence 5, Appli
17	41	41.4	10	2	US-08-985-126-13	Sequence 13, Appl
18	41	41.4	10	5	PCT-US95-11126-5	Sequence 5, Appli
19	41	41.4	10	5	PCT-US95-11126-13	Sequence 13, Appl
20	40	40.4	10	1	US-08-299-285-14	Sequence 14, Appl
21	40	40.4	10	1	US-08-299-285-23	Sequence 23, Appl
22	40	40.4	10	2	US-08-985-126-14	Sequence 14, Appl
23	40	40.4	10	2	US-08-985-126-23	Sequence 23, Appl
24	40	40.4	10	5	PCT-US95-11126-14	Sequence 14, Appl
25	40	40.4	10	5	PCT-US95-11126-23	Sequence 23, Appl
26	39	39.4	506	4	US-08-942-012B-26	Sequence 26, Appl
27	38	38.4	19	1	US-08-191-337-4	Sequence 4, Appli

Sequence 25, Appli
Patent No. 5180581
Sequence 5, Appli
Sequence 58, Appli
Sequence 48, Appli
Sequence 52, Appli
Sequence 60, Appli
Sequence 50, Appli
Sequence 54, Appli
Sequence 61, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 12, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 73, Appli
Sequence 74, Appli
Sequence 71, Appli

28 38 38.4 506 4 US-08-942-012B-25
29 38 38.4 506 6 5180581-2
30 38 38.4 539 3 US-08-557-210A-5
31 37.5 37.9 215 2 US-08-912-129A-58
32 37.5 37.9 245 2 US-08-912-129A-48
33 37.5 37.9 373 2 US-08-912-129A-52
34 37.5 37.9 460 2 US-08-912-129A-60
35 37.5 37.9 490 2 US-08-912-129A-50
36 37.5 37.9 618 2 US-08-912-129A-54
37 37.5 37.9 873 2 US-08-912-129A-61
38 37 37.4 10 1 US-08-299-285-12
39 37 37.4 10 2 US-08-985-126-12
40 37 37.4 10 5 PCT-US95-11126-12
41 37 37.4 199 2 US-08-405-175A-3
42 37 37.4 200 2 US-08-405-175A-4
43 37 37.4 395 3 US-08-781-891-73
44 37 37.4 1269 3 US-08-781-891-74
45 37 37.4 1432 3 US-08-781-891-71

ALIGNMENTS

RESULT 1
US-08-463-224-3
; Sequence 3, Application US/08463224
; Patent No. 5807824
; GENERAL INFORMATION:
; APPLICANT: van Oostrum, Jan
; APPLICANT: Boyar, William C.
; APPLICANT: Galakatos, Nicholas G.
; APPLICANT: Schmitz, Albert
; APPLICANT: van Heeke, Gino
; TITLE OF INVENTION: C5a Receptor Antagonists Having
; TITLE OF INVENTION: Substantially No. 5807824Agonist Activity
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
; STREET: 600 South Avenue West
; CITY: Westfield
; STATE: NJ
; COUNTRY: USA
; ZIP: 07090
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,224
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Foley, Shawn P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 908-654-5000
; TELEFAX: 908-654-7866
; TELEX: 139-125
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 74 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-463-224-3

Query Match 47.5% Score 47: DB 1: Length 74:
Best Local Similarity 100.0%: Pred. No. 0.35:
Matches 9: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

QY 1 KYKHSVKK 9
|||||||

```
Db      12 KYKHSVVK 20

RESULT 2
US-08-463-377-3
: Sequence 3, Application US/08463377
: Patent No. 5837499
: GENERAL INFORMATION:
: APPLICANT: van Oostrum, Jan
: APPLICANT: Boyar, William C.
: APPLICANT: Galakatos, Nicholas G.
: APPLICANT: Schmitz, Albert
: APPLICANT: van Heeke, Gino
: TITLE OF INVENTION: C5a Receptor Antagonists Having
: TITLE OF INVENTION: Substantially No. 5837499Agonist Activity
: NUMBER OF SEQUENCES: 67
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lerner, David, Littenberg, Krumholz & Mentlik
: STREET: 600 South Avenue West
: CITY: Westfield
: STATE: NJ
: COUNTRY: USA
: ZIP: 07090
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/463,377
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Foley, Shawn P.
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 908-654-5000
: TELEFAX: 908-654-7866
: TELEX: 139-125
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 74 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-463-377-3

Query Match      47.5%; Score 47; DB 2; Length 74;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYKHSVVK 9
       11111111
Db      12 KYKHSVVK 20

RESULT 3
US-09-246-500B-10
: Sequence 10, Application US/09246500B
: Patent No. 6235494
: GENERAL INFORMATION:
: APPLICANT: Hugli, Tony E.
: TITLE OF INVENTION: Substrates for Assessing Mannan-Binding
: TITLE OF INVENTION: Protein-Associated Serine Protease Activity and Methods
: TITLE OF INVENTION: Using the Substrates
: FILE REFERENCE: 24730-2204
: CURRENT APPLICATION NUMBER: US/09/246,500B
: CURRENT FILING DATE: 1999-02-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 10
: LENGTH: 74
: TYPE: PRT

ORGANISM: Human C5a Anaphylatoxin
FEATURE:
NAME/KEY: CARBOHYD
LOCATION: (64)...(66)
US-09-246-500B-10

Query Match      47.5%; Score 47; DB 4; Length 74;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYKHSVVK 9
       11111111
Db      12 KYKHSVVK 20

RESULT 4
US-09-246-500B-14
: Sequence 14, Application US/09246500B
: Patent No. 6235494
: GENERAL INFORMATION:
: APPLICANT: Hugli, Tony E.
: TITLE OF INVENTION: Substrates for Assessing Mannan-Binding
: TITLE OF INVENTION: Protein-Associated Serine Protease Activity and Methods
: TITLE OF INVENTION: Using the Substrates
: FILE REFERENCE: 24730-2204
: CURRENT APPLICATION NUMBER: US/09/246,500B
: CURRENT FILING DATE: 1999-02-08
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 14
: LENGTH: 77
: TYPE: PRT
: ORGANISM: Mouse C5a Anaphylatoxin
: US-09-246-500B-14

Query Match      47.5%; Score 47; DB 4; Length 77;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 KYKHSVVK 9
       11111111
Db      15 KYKHSVVK 23

RESULT 5
US-08-299-285-7
: Sequence 7, Application US/08299285
: Patent No. 5696230
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: TITLE OF INVENTION: C-terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,285
: FILING DATE: 31-AUG-1994
```

CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-299-285-7

Query Match 42.4%; Score 42; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 YSFKPMPLXR 20
IIII IIII
DB 1 YSFKDMPLAR 10

RESULT 6
US-08-299-285-10
Sequence 10, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO

ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: "D-Alanine at position 9"
US-08-299-285-10

Query Match 42.4%; Score 42; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 YSFKPMPLXR 20
IIII IIII
DB 1 YSFKDMPLAR 10

RESULT 7
US-08-299-285-24
Sequence 24, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: "D-Alanine at position 9"
US-08-299-285-24

Query Match 42.4%; Score 42; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;


```
QY 11 YSEKPMPLXR 20
      ||||| I I
Db 1 YSEKPMOLAR 10

RESULT 8
US-08-985-126-7
: Sequence 7, Application US/08985126
: Patent No. 5942599
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/985,126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,285
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: No. 5942599 Relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: "D-Alanine at position 9"
: US-08-985-126-10

Query Match 42.4%; Score 42; DB 2; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 11 YSEKPMPLXR 20
      ||||| I I
Db 1 YSEKPMPLAR 10

RESULT 10
US-08-985-126-24
: Sequence 24, Application US/08985126
: Patent No. 5942599
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
```

```

STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: No. 5942599 Relevant
TOPOLOGY: No. 5942599 Relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: "D-Alanine at position 9"
US-08-985-126-24

Query Match 42.4%; Score 42; DB 2; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 11 YSEKPMPLXR 20
    | | | | | | | |
Db 1 YSEKPMQLAR 10

RESULT 11
PCT-US95-11126-7
: Sequence 7, Application PC/TUS9511126
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnatsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: TITLE OF INVENTION: C-terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/299,285
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:

```

```

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
PCT-US95-11126-7

Query Match 42.4%; Score 42; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY 11 YSEKPMPLXR 20
    | | | | | | | |
Db 1 YSEKDMPLAR 10

RESULT 12
PCT-US95-11126-10
: Sequence 10, Application PC/TUS9511126
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnatsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: TITLE OF INVENTION: C-terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/11126
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/299,285
: FILING DATE: 31-AUG-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:

```

LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 9
OTHER INFORMATION: /product= "D-Alanine at position 9"
PCT-US95-11126-10

Query Match 42.4%; Score 42; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 11 YSFKPMPPLXR 20
|||||
Db 1 YSFKDMPPLAR 10

RESULT 13

PCT-US95-11126-24
Sequence 24, Application PC/TUS9511126
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman, P.C.
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/11126
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/299,285
FILING DATE: 31-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site

LOCATION: 9
OTHER INFORMATION: /product= "D-Alanine at position 9"
PCT-US95-11126-24

Query Match 42.4%; Score 42; DB 5; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.32;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 11 YSFKPMPPLXR 20
|||||
Db 1 YSFKPMPPLAR 10

RESULT 14

US-08-299-285-5
Sequence 5, Application US/08299285
Patent No. 5696230
GENERAL INFORMATION:
APPLICANT: Sanderson, Sam D.
APPLICANT: Sherman, Simon A.
APPLICANT: Kirnarsky, Leonid
APPLICANT: Taylor, Stephen M.
TITLE OF INVENTION: High Affinity Response-Selective
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,285
FILING DATE: 31-AUG-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
REFERENCE/DOCKET NUMBER: 63075
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
US-08-299-285-5

Query Match 41.4%; Score 41; DB 1; Length 10;
Best Local Similarity 80.0%; Pred. No. 0.47;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 11 YSFKPMPPLXR 20
|||||
Db 1 YSFKDMPPLGR 10

RESULT 15

US-08-299-285-13

: Sequence 13, Application US/08299285
: Patent No. 5696230
: GENERAL INFORMATION:
: APPLICANT: Sanderson, Sam D.
: APPLICANT: Sherman, Simon A.
: APPLICANT: Kirnarsky, Leonid
: APPLICANT: Taylor, Stephen M.
: TITLE OF INVENTION: High Affinity Response-Selective
: TITLE OF INVENTION: C-Terminal Analogs of C5a Anaphylatoxin
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: US
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/299,285
: FILING DATE: 31-AUG-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: REFERENCE/DOCKET NUMBER: 63075
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: FEATURE:
: NAME/KEY: Modified-site
: LOCATION: 9
: OTHER INFORMATION: "D-Phenylalanine at
: OTHER INFORMATION: position 9"
: US-08-299-285-13

Query Match 41.4% Score 41; DB 1; Length 10;
Best Local Similarity 80.0% Pred. No. 0.47;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Caps 0;
Oy 11 YSFKDPLXR 20
11111111
Db 1 YSFKDPLXR 10

Search completed: February 27, 2002, 11:36:03
Job time: 142 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:41:14 ; Search time 303.5 Seconds
(without alignments)
1.464 Million cell updates

Title: US-09-446-109A-7
Perfect score: 35
Sequence: 1 FKXPWR 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101:★

1:	/SID58/gcgdata/geneseq/geneseqp/AA1980.DAT: *
2:	/SID58/gcgdata/geneseq/geneseqp/AA1981.DAT: *
3:	/SID58/gcgdata/geneseq/geneseqp/AA1982.DAT: *
4:	/SID58/gcgdata/geneseq/geneseqp/AA1983.DAT: *
5:	/SID58/gcgdata/geneseq/geneseqp/AA1984.DAT: *
6:	/SID58/gcgdata/geneseq/geneseqp/AA1985.DAT: *
7:	/SID58/gcgdata/geneseq/geneseqp/AA1986.DAT: *
8:	/SID58/gcgdata/geneseq/geneseqp/AA1987.DAT: *
9:	/SID58/gcgdata/geneseq/geneseqp/AA1988.DAT: *
10:	/SID58/gcgdata/geneseq/geneseqp/AA1989.DAT: *
11:	/SID58/gcgdata/geneseq/geneseqp/AA1990.DAT: *
12:	/SID58/gcgdata/geneseq/geneseqp/AA1991.DAT: *
13:	/SID58/gcgdata/geneseq/geneseqp/AA1992.DAT: *
14:	/SID58/gcgdata/geneseq/geneseqp/AA1993.DAT: *
15:	/SID58/gcgdata/geneseq/geneseqp/AA1994.DAT: *
16:	/SID58/gcgdata/geneseq/geneseqp/AA1995.DAT: *
17:	/SID58/gcgdata/geneseq/geneseqp/AA1996.DAT: *
18:	/SID58/gcgdata/geneseq/geneseqp/AA1997.DAT: *
19:	/SID58/gcgdata/geneseq/geneseqp/AA1998.DAT: *
20:	/SID58/gcgdata/geneseq/geneseqp/AA1999.DAT: *
21:	/SID58/gcgdata/geneseq/geneseqp/AA2000.DAT: *
22:	/SID58/gcgdata/geneseq/geneseqp/AA2001.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	6	13 AAR30577	Anaphylatoxin anta
2	33	94.3	6	16 AAR89737	C5a peptide analog
3	30	85.7	39	20 AAW74031	Human D2H binding
4	30	85.7	434	21 AAB01208	Corn putative leci
5	30	85.7	462	22 AAG36875	E. coli growth and
6	30	85.7	700	21 AAY58359	Yeast G protein-co
7	30	85.7	847	21 AAY58360	Yeast G protein-co
8	30	85.7	864	21 AAY58358	Yeast G protein-co
9	30	85.7	953	21 AAY58356	Yeast G protein-co
10	30	85.7	953	21 AAY58357	Yeast G protein-co
11	30	85.7	961	21 AAY58355	Yeast G protein-co

ALIGNMENTS

RESULT	1
AAR30577	
ID	AAR30577 standard; peptide; 6 AA.
XX	
AC	
XX	AAR30577;
DT	
XX	29-JAN-1993 (first entry)
DE	Anaphylatoxin antagonist or agonist peptide.
XX	
KW	Anaphylatoxin; receptor; ligand; antiinflammatory; hexapeptide;
KW	heptapeptide; immunodeficiency; allergy; autoimmune; ARDS; cancer;
KW	infection; endotoxin; asthma; gout; psoriasis; cirrhosis;
KW	inflammatory; bowel; disease; hepatitis; burns; myocardial;
KW	infarction; transplant rejection; ischaemic.
XX	
OS	Synthetic.
XX	
FH	Key
FT	Modified-site
FT	Location/Qualifiers
FT	1
FT	/note= "N-Me-phe"
FT	4
FT	/label= OTHER
FT	/note= "(2R)-2-amino-3-cyclohexyl-propanoyl"
FT	6
FT	/note= "D-Arg"
XX	
XX	WO9211858-A.
PN	
PD	23-JUL-1992.
XX	
PF	10-DEC-1991;
XX	91WO-US09319.
PR	27-DEC-1990;
XX	90US-0634641.

XX (ABBO) ABBOTT LABORATORIES.
 XX Kawai M, Luly JR, Or YS, Wagner R, Wiedeman PE;
 PI WPI; 1992-268383/32.
 XX
 XX New hexa- and hepta-peptide(s) are anaphylatoxin antagonists and
 PT agonists - for treating inflammatory and immunodeficiency
 PT diseases, cancers and severe infections
 XX
 XX Claim 11; Page 153; 160pp; English.
 PS
 XX The peptide is a specifically claimed example of a group of highly
 CC generic hexa- and heptapeptides which are (a) anaphylatoxin
 CC antagonists useful for treating asthma, other allergies,
 CC inflammations, autoimmune diseases, serum sickness, gout, bullous
 CC skin diseases, psoriasis, ARDS, endotoxin shock, hepatic cirrhosis,
 CC pancreatitis, inflammatory bowel disease, burns, sepsis, myocardial
 CC infarction, chronic hepatitis, transplant rejection, or ischaemic
 CC heart or brain damage; or (b) anaphylatoxin agonists useful for
 CC stimulating inflammatory and immune responses, e.g. in the treatment
 CC of cancer, immunodeficiency diseases and severe infections.
 XX
 XX Sequence 6 AA:
 SQ

Query Match 94.3%; Score 33; DB 13; Length 6;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPxWR 6
 Db 1 fkpwxr 6
 |||||

RESULT 2
 AAR89737
 ID AAR89737 standard; peptide: 6 AA.
 XX
 AC AAR89737;
 XX
 DT 03-MAY-1996 (first entry)
 XX
 DE C5a peptide analogue, C089, used to identify C5a (ant)agonists.
 XX
 KW C5a; complement; agonist; antagonist; peptide; binding assay;
 KW identification; inflammation; pain reduction; respiratory disorder;
 KW cardiovascular; hypertension; ARDS; anaphylaxis; arthritis; asthma;
 KW competitive.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 1
 FT /label= Phe, OTHER
 FT /note= "OTHER - the Bolton-Hunter modified peptide
 FT having the 3-(p-hydroxyphenyl)-propionyl
 FT group"
 FT
 FT Misc-difference 4
 FT /label= OTHER
 FT /note= "D-cyclohexylalanine"
 FT
 FT Modified-site 6
 FT /label= OTHER
 FT /note= "cyclohexylalanine"
 FT
 FT
 XX W09525957-A1.
 XX
 PD 28-SEP-1995.
 XX
 XX 14-MAR-1995; 95WO-US03209.
 XX
 XX 18-MAR-1994; 94US-0215137.

XX (MERI) MERCK & CO INC.
 XX Kondeatis 2, Siciliano SJ, Springer MS;
 PI WPI; 1995-344718/44.
 XX
 XX C5a receptor binding assays - used for identifying cpds. having C5a
 PT antagonist, agonist or partial agonist activity
 XX
 XX Claim 15; Page 47; 65pp; English.
 PS
 XX C5a C-terminal peptide analogues AAR89734-44 and AAR90033-35 may be used
 CC in a new method for identifying C5a agonists/antagonists. AAR89734
 CC represents the generic formula of this newly defined class of
 CC labelled peptides. C5a antagonists identified using this method are
 CC useful in the treatment of a wide variety of C5a-mediated diseases
 CC such as acute respiratory distress syndrome (ARDS), anaphylactic
 CC shock, psoriasis, osteoarthritis, rheumatoid arthritis, asthma,
 CC cardiovascular disorders, and metastatic spread of cancerous tumours.
 XX
 XX Sequence 6 AA:
 SQ

Query Match 94.3%; Score 33; DB 16; Length 6;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPxWR 6
 Db 1 fkpwxr 6
 |||||

RESULT 3
 AAW74031
 ID AAW74031 standard; peptide: 39 AA.
 XX
 AC AAW74031;
 XX
 DT 04-MAY-1999 (first entry)
 XX
 DE Human D2H binding protein DAB3.
 XX
 KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
 KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 KW therapeutic agent delivery; therapy.
 XX
 OS Homo sapiens.
 XX
 FN W09851325-A2.
 XX
 PD 19-NOV-1998.
 XX
 IF 15-MAY-1998; 98WO-US10088.
 XX
 PR 15-MAY-1997; 97US-0046595.
 XX
 XX (CYTO-) CYTOGEN CORP.
 XX (ELAN-) ELAN CORP PLC.
 XX
 PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin TJ;
 PI Omahony DJ, Patterson CA, Singleton J;
 XX
 DR WPI; 1999-009568/01.
 XX
 XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimerae and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX
 PS Claim 2; Page 54; 294pp; English.

XX This sequence represents a peptide that specifically binds to the human
 CC D2H protein. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid. Imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 XX Sequence 39 AA;

Query Match 85.7%; Score 30; DB 20; Length 39;
 Best Local Similarity 66.7%; Pred. No. 18;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 I: I I I
 DB 15 frpgwr 20

RESULT 4
 AAB01208
 ID AAB01208 standard; Protein; 434 AA.

XX AAB01208;

XX 12-DEC-2000 (first entry)

DE Corn putative lecithin:cholesterol acyltransferase #3.

XX Corn; lecithin:cholesterol acyltransferase; phytosterol;
 KW phosphatidylcholine-sterol O-acyltransferase; heat shock; cold shock.
 XX Zea mays.

XX WO200032791-A2.

XX 08-JUN-2000.

XX 02-DEC-1999; 99WO-US28586.

XX 03-DEC-1998; 98US-0110782.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Cahoon RE, Kinney AJ, Sakai H, Shen JB, Butler KH, Saylor JU;

DR WPI: 2000-412337/35.

DR N-PSDB; AAA49203.

XX Polynucleotide encoding plant lecithin:cholesterol acyltransferase
 CC enzyme useful for producing transgenic plants and for producing
 PT antibodies specific to which is useful for screening cDNA expression
 PT libraries

XX Claim 10; Page 40-41; 49pp; English.

XX The present sequence is a putative protein sequence of a corn
 CC lecithin:cholesterol acyltransferase (also known as
 CC phosphatidylcholine-sterol O-acyltransferase). This enzyme is found
 CC associated with high-density lipoproteins and Apolipoprotein-A1 and -B.

CC The gene and protein can be used to produce transgenic plants which have
 CC increased lipid metabolism and membrane fluidity, and therefore increased
 CC resistance to heat and/or cold shock, to alter the content of phytosterol
 CC or lecithin in grains and to identify potential herbicides.

XX Sequence 434 AA;

Query Match 85.7%; Score 30; DB 21; Length 434;
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 I: I I I
 DB 271 frpmwr 276

RESULT 5
 AAG98875
 ID AAG98875 standard; Protein; 462 AA.

XX AAG98875;

XX 26-SEP-2001 (first entry)

DE E. coli growth and proliferation related protein sequence SEQ ID NO:345.

XX Escherichia coli; growth; proliferation; microbial; antimicrobial;
 KW bacterial infection; microorganism.

XX Escherichia coli.

XX WO200134810-A2.

XX 17-MAY-2001

XX 09-NOV-2000; 2000WO-US30950.

XX 09-NOV-1999; 99US-0164415.

XX (ELIT-) ELITRA PHARM INC.

XX Forsyth RA, Ohlsen K, Zyskind J;

XX WPI: 2001-335933/35.

XX N-PSDB; AAH84546.

XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful
 FT for screening for homologous genes and for designing expression vectors

XX Claim 19; Page 421-422; 522pp; English.

XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
 CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
 CC growth and proliferation related proteins given in AAG99078 and AAG98830
 CC to AAG99999. (I) can be used as potential targets for the generation of
 CC new antimicrobial agents, and for identification of compounds which
 CC interact with the gene products of (I). In addition the expression of
 CC (I) and the purification of the proteins, the purified proteins can be
 CC used to generate reagents and screen small molecule libraries or other
 CC candidate compound libraries for compounds that can be further developed
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
 CC complementary to (I) that are specific for particular species of
 CC microorganisms can be used to identify particular microorganism species
 CC in clinical specimens, therefore, providing a rapid and dependable
 CC method by which to identify the causative agents of a bacterial
 CC infection. Also, antibodies generated against proteins translated from
 CC mRNA transcribed from proliferation-required sequences can also be used
 CC to screen for specific microorganisms that produce such proteins in a
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing
 CC primers used in the isolation of E. coli growth and proliferation
 CC related sequence, which are used in an example from the present


```
CC invention.
XX
SQ Sequence 462 AA;

Query Match 85.7%; Score 30; DB 22; Length 462;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
   |||
Db 398 ykplwr 403

RESULT 6
AA58359
ID AAY58359 standard; Protein; 700 AA.
XX
AC AAY58359;
XX
DT 27-MAR-2000 (first entry)
XX
DE Yeast G protein-coupled receptor GPR1 mutant d694-954.
XX
KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
KW pseudohyphal form; signalling pathway; antifungal; mutant; mutein.
XX
OS Saccharomyces cerevisiae.
XX Synthetic.
XX
PN W09963094-A2.
XX
PD 09-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11838.
XX
PR 01-JUN-1998; 98US-0088311.
XX
PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
PI Hirsch JP, Xue Y;
XX
DR WPI: 2000-086980/07.
XX
PT Novel protein, gene, antibody and recombinant cell useful for
PT identifying modulators and inhibitors useful as anti-fungals -
XX
PS Example 6.2.5; Page -: 63pp; English.
XX
CC The invention relates to a novel yeast G protein-coupled receptor,
CC GPR1 (AAY58355). GPR1 is predicted to have seven transmembrane domains,
CC which is characteristic of G protein-coupled receptors, and also
CC contains a very large third cytoplasmic loop and a large cytoplasmic
CC tail. The third cytoplasmic loop contains two short basic sequences; one
CC sequence (KRKAQIC) is present at the N-terminal end of the loop and the
CC other sequence (KKRAQIC) is at the C-terminal end of the loop. The
CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
CC believed to provide the upstream signal that activates GPA2, and plays
CC an essential role in inducing the switch from non-filamentous to
CC filamentous growth in yeast. Modulators of GPR1 function, especially
CC inhibitors, are useful for prevention or treatment of a disease or
CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
CC inserted heterologous sequence inactivates the biological activity of
CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
CC activity may be used to inhibit the conversion of a fungus from a
CC non-filamentous form to a filamentous form. The present sequence
CC represents a GPR mutant, d277-284, used in an exemplification of
CC the invention. This sequence is lacking the majority of the cytoplasmic
CC C-terminal tail, relative to the native GPR1.
CC Note: This sequence is not shown in the specification, but is derived
CC from the yeast GPR1 amino acid sequence given in figure 1A.
XX
```

```
SQ Sequence 700 AA;

Query Match 85.7%; Score 30; DB 21; Length 700;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
   |||
Db 156 fkpwnk 161

RESULT 7
AA58360
ID AAY58360 standard; Protein; 847 AA.
XX
AC AAY58360;
XX
DT 27-MAR-2000 (first entry)
XX
DE Yeast G protein-coupled receptor GPR1 mutant d841-954.
XX
KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
KW pseudohyphal form; signalling pathway; antifungal; mutant; mutein.
XX
OS Saccharomyces cerevisiae.
XX Synthetic.
XX
PN W09963094-A2.
XX
PD 09-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11838.
XX
PR 01-JUN-1998; 98US-0088311.
XX
PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
PI Hirsch JP, Xue Y;
XX
DR WPI: 2000-086980/07.
XX
PT Novel protein, gene, antibody and recombinant cell useful for
PT identifying modulators and inhibitors useful as anti-fungals -
XX
PS Example 6.2.5; Page -: 63pp; English.
XX
CC The invention relates to a novel yeast G protein-coupled receptor,
CC GPR1 (AAY58355). GPR1 is predicted to have seven transmembrane domains,
CC which is characteristic of G protein-coupled receptors, and also
CC contains a very large third cytoplasmic loop and a large cytoplasmic
CC tail. The third cytoplasmic loop contains two short basic sequences; one
CC sequence (KRKAQIC) is present at the N-terminal end of the loop and the
CC other sequence (KKRAQIC) is at the C-terminal end of the loop. The
CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
CC believed to provide the upstream signal that activates GPA2, and plays
CC an essential role in inducing the switch from non-filamentous to
CC filamentous growth in yeast. Modulators of GPR1 function, especially
CC inhibitors, are useful for prevention or treatment of a disease or
CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
CC inserted heterologous sequence inactivates the biological activity of
CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
CC activity may be used to inhibit the conversion of a fungus from a
CC non-filamentous form to a filamentous form. The present sequence
CC represents a GPR mutant, d277-284, used in an exemplification of
CC the invention. This sequence is lacking a portion (designated the
CC Smaest region) of the cytoplasmic C-terminal tail, relative to the
CC native GPR1.
CC Note: This sequence is not shown in the specification, but is derived
CC from the yeast GPR1 amino acid sequence given in figure 1A.
XX
SQ Sequence 847 AA;
```

Query Match 85.7%; Score 30; DB 21; Length 847;
 Best Local Similarity 66.7%; Pred. No. 4.2e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 6
 III I;
 DB 156 fknwk 161

RESULT 8
 ID AAY58358 standard; Protein; 864 AA.
 XX AC AAY58358;
 XX DT 27-MAR-2000 (first entry)
 XX DE Yeast G protein-coupled receptor GPR1 mutant d490-586.
 XX KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
 XX KW pseudohyphal form; signalling pathway; antifungal; mutant; mutein.
 XX OS Saccharomyces cerevisiae.
 XX OS Synthetic.
 XX PN W09963094-A2.
 XX PD 09-DEC-1999.
 XX PD 09-DEC-1999.
 XX PF 28-MAY-1999; 99WO-US11838.
 XX PF 28-MAY-1999; 99WO-US11838.
 XX PR 01-JUN-1998; 98US-0088311.
 XX PR 01-JUN-1998; 98US-0088311.
 XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.
 XX PI Hirsch JP, Xue Y;
 XX PT WPT; 2000-086980/07.
 XX PT Novel protein, gene, antibody and recombinant cell useful for
 PT identifying modulators and inhibitors useful as anti-fungals -
 PS Example 6.2.5; Page -: 63pp; English.

XX The invention relates to a novel yeast G protein-coupled receptor,
 CC GPR1 (AAY58358). GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences: one
 CC sequence (KKRIKAQIG) is present at the N-terminal end of the loop and the
 CC other sequence (KKRAQIQ) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form. The present sequence of
 CC represents a GPR mutant, d277-284, used in an exemplification of
 CC the invention. This sequence is lacking the asparagine-rich region
 CC of the third cytoplasmic loop, relative to the native GPR1.
 CC Note: This sequence is not shown in the specification, but is derived
 CC from the yeast GPR1 amino acid sequence given in figure 1A.

XX Sequence 864 AA:
 SQ

Query Match 85.7%; Score 30; DB 21; Length 953;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;

QY 1 FKXWR 6
 III I;
 DB 156 fknwk 161

RESULT 9
 ID AAY58356 standard; Protein; 953 AA.
 XX AC AAY58356;
 XX DT 27-MAR-2000 (first entry)
 XX DE Yeast G protein-coupled receptor GPR1 mutant d277-284.
 XX KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
 XX KW pseudohyphal form; signalling pathway; antifungal; mutant; mutein.
 XX OS Saccharomyces cerevisiae.
 XX OS Synthetic.
 XX PN W09963094-A2.
 XX PD 09-DEC-1999.
 XX PD 09-DEC-1999.
 XX PF 28-MAY-1999; 99WO-US11838.
 XX PF 28-MAY-1999; 99WO-US11838.
 XX PR 01-JUN-1998; 98US-0088311.
 XX PR 01-JUN-1998; 98US-0088311.
 XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.
 XX PI Hirsch JP, Xue Y;
 XX PT WPT; 2000-086980/07.
 XX PT Novel protein, gene, antibody and recombinant cell useful for
 PT identifying modulators and inhibitors useful as anti-fungals -
 PS Example 6.2.5; Page -: 63pp; English.

XX The invention relates to a novel yeast G protein-coupled receptor,
 CC GPR1 (AAY58356). GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences: one
 CC sequence (KKRIKAQIG) is present at the N-terminal end of the loop and the
 CC other sequence (KKRAQIQ) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form. The present sequence of
 CC represents a GPR mutant, d277-284, used in an exemplification of
 CC the invention. This sequence is lacking the asparagine-rich region
 CC of the third cytoplasmic loop, relative to the native GPR1.
 CC Note: This sequence is not shown in the specification, but is derived
 CC from the yeast GPR1 amino acid sequence given in figure 1A.

XX Sequence 953 AA:
 SQ

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
1111:
Db 156 fkpwnk 161

RESULT 10
AA58357
ID AAY58357 standard; Protein: 953 AA.

XX AC AAY58357;
XX DT 27-MAR-2000 (first entry)

XX DE Yeast G protein-coupled receptor GPR1 mutant d610-617.

XX KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
XX KW pseudohyphal form; signalling pathway; antifungal; mutant; muten.

XX OS Saccharomyces cerevisiae.
XX OS Synthetic.

XX PN WO9963094-A2.

XX PD 09-DEC-1999.

XX PF 28-MAY-1999; 99WO-US11838.

XX PR 01-JUN-1998; 98US-0088311.

XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.

XX PI Hirsch JP, Xue Y;

XX DR WPI; 2000-086980/07.

XX PT Novel protein, gene, antibody and recombinant cell useful for
XX PT identifying modulators and inhibitors useful as anti-fungals -

XX PS Example 6.2.5; Page -: 63pp; English.

XX CC The invention relates to a novel yeast G protein-coupled receptor,
XX CC GPR1 (AAY58355). GPR1 is predicted to have seven transmembrane domains,
XX CC which is characteristic of G protein-coupled receptors, and also
XX CC contains a very large third cytoplasmic loop and a large cytoplasmic
XX CC tail. The third cytoplasmic loop contains two short basic sequences: one
XX CC other sequence (KRKAQIQ) is present at the N-terminal end of the loop and the
XX CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
XX CC believed to provide the upstream signal that activates GPA2, and plays
XX CC an essential role in inducing the switch from non-filamentous to
XX CC filamentous growth in yeast. Modulators of GPR1 function, especially
XX CC inhibitors, are useful for prevention or treatment of a disease or
XX CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
XX CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
XX CC inserted heterologous sequence inactivates the biological activity of
XX CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
XX CC activity may be used to inhibit the conversion of a fungus from a
XX CC non-filamentous form to a filamentous form. The present sequence
XX CC represents a GPR mutant, d610-617, used in an exemplification of
XX CC the invention. This sequence is lacking the second basic motif
XX CC (KKRRAQIQ) of the third cytoplasmic loop, relative to the native GPR1.
XX CC Note: This sequence is not shown in the specification, but is derived
XX CC from the yeast GPR1 amino acid sequence given in figure 1A.

XX SQ Sequence 953 AA;

Query Match 85.7%; Score 30; DB 21; Length 953;
Best Local Similarity 66.7%; Pred. No. 4.7e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
1111:
Db 156 fkpwnk 161

RESULT 11
AAY58355
ID AAY58355 standard; Protein: 961 AA.

XX AC AAY58355;
XX DT 27-MAR-2000 (first entry)
XX DE Yeast G protein-coupled receptor GPR1.
XX KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
XX KW pseudohyphal form; signalling pathway; antifungal.
XX OS Saccharomyces cerevisiae.

XX FH Key Location/Qualifiers
XX FT Region 1..55
XX FT /note= "Extracellular N-terminus"
XX FT Domain 56..80
XX FT /note= "Transmembrane domain 1"
XX FT Region 81..90
XX FT /note= "Intracellular loop 1"
XX FT Domain 91..115
XX FT /note= "Transmembrane domain 2"
XX FT Region 116..132
XX FT /note= "Extracellular loop 1"
XX FT Domain 133..156
XX FT /note= "Transmembrane domain 3"
XX FT Region 157..180
XX FT /note= "Intracellular loop 2"
XX FT Domain 181..200
XX FT /note= "Transmembrane domain 4"
XX FT Region 201..246
XX FT /note= "Extracellular loop 2"
XX FT Domain 247..276
XX FT /note= "Transmembrane domain 5"
XX FT Misc-difference 250
XX FT /note= "Encoded by AGC"
XX FT Region 277..620
XX FT /note= "Intracellular loop 3"
XX FT Region 277..284
XX FT /note= "Basic motif 1 (KRKAQIQ)"
XX FT Region 490..586
XX FT /note= "Asparagine-rich region"
XX FT Region 610..617
XX FT /note= "Basic motif 2 (KKRRAQIQ)"
XX FT Domain 621..645
XX FT /note= "Transmembrane domain 6"
XX FT Region 646..656
XX FT /note= "Extracellular loop 3"
XX FT Domain 657..678
XX FT /note= "Transmembrane domain 7"
XX FT Region 679..961
XX FT /note= "Intracellular C-terminus"
XX FT Misc-difference 250
XX FT /note= "Encoded by AGC"

XX PN WO9963094-A2.
XX PD 09-DEC-1999.
XX PF 28-MAY-1999; 99WO-US11838.
XX PR 01-JUN-1998; 98US-0088311.
XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.
XX PI Hirsch JP, Xue Y;

XX
DR WPI: 2000-086980/07.
DR N-PSDB; AA255699.
XX
PT Novel protein, gene, antibody and recombinant cell useful for
PT identifying modulators and inhibitors useful as anti-fungals -
XX
PS Claim 4: Fig 1A; 63pp; English.
XX
CC This sequence represents a novel yeast G protein-coupled receptor,
CC GPRI. GPRI is predicted to have seven transmembrane domains,
CC which is characteristic of G protein-coupled receptors, and also
CC contains a very large third cytoplasmic loop and a large cytoplasmic
CC tail. The third cytoplasmic loop contains two short basic sequences; one
CC sequence (KKRAQIG) is present at the N-terminal end of the loop and the
CC other sequence (KKRAQIQ) is at the C-terminal end of the loop. The
CC third cytoplasmic loop also contains an asparagine-rich region. GPRI is
CC believed to provide the upstream signal that activates GPA2, and plays
CC an essential role in inducing the switch from non-filamentous to
CC filamentous growth in yeast. Modulators of GPRI function, especially
CC inhibitors, are useful for prevention or treatment of a disease or
CC disorder involving a fungal infection. The inhibitor may be an anti-GPRI
CC antibody, GPRI antisense nucleic acid or a GPRI gene altered so that an
CC inserted heterologous sequence inactivates the biological activity of
CC GPRI. Compounds which modulate GPRI gene expression and/or gene product
CC activity may be used to inhibit the conversion of a fungus from a
CC non-filamentous form to a filamentous form.
XX
XX
SQ Sequence 961 AA:

Query Match 85.7%; Score 30; DB 21; Length 961;
Best Local Similarity 66.7%; Pred. NO. 4.8e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPXWR 6
Db 156 fkpww 161

RESULT 12
AAR41001
ID AAR41001 standard; Protein; 1093 AA.

XX
AC AAR41001;
XX
DT 25-FEB-1994 (first entry)
XX
DE Human myotonic dystrophy gene' protein.
XX
KW Abnormality: muscular dystrophy; CHR 19; chromosome 19;
KW protein kinase; polymerase chain reaction; brain.
XX
OS Homo sapiens.

XX
FH Key Location/Qualifiers
FT Region 1..1093
FT /note= "encoded by predicted reading frame a,
FT x's in the sequence indicate stop codons
FT in the reading frame"

XX
PN W09317104-A.
XX
PD 02-SEP-1993.
XX
PF 19-FEB-1993; 93WO-US01545.
XX
PR 20-FEB-1992; 92US-0839255.
XX
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Brook JD, Housman DE;

DR WPI: 1993-288410/36.
XX
PT DNA sequence of myotonic dystrophy gene - used to produce probes
PT and identify CHR 19 abnormality and protein kinase responsible
XX
PS Disclosure; Fig 6; 64pp; English.
XX
CC The sequence is that encoded by predicted reading frame a of
CC the human myotonic dystrophy (DM) gene. It may be used in the
CC identification of individuals affected by DM.
XX
SQ Sequence 1093 AA:

Query Match 85.7%; Score 30; DB 14; Length 1093;
Best Local Similarity 66.7%; Pred. NO. 5.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPXWR 6
Db 382 frprwr 387

RESULT 13
AAV51233
ID AAV51233 standard; Protein; 2204 AA.

XX
AC AAV51233;
XX
DT 07-APR-2000 (first entry)
XX
DE Newcastle disease virus LaSota genome encoded protein 6.
XX
KW Avian-paramyxovirus; infection; lentogenic; F protein; vaccine;
KW respiratory disease; gastrointestinal disease; poultry pathogen;
KW local immunity.
XX
OS Newcastle disease virus.

XX
PN W09966045-A1.

XX
PD 23-DEC-1999.
XX
PF 17-JUN-1999; 99WO-NL00377.
XX
PR 19-JUN-1998; 98EP-0202054.

XX
PA (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.

XX
PI Peeters BPH, De Leeuw OS, Koch G, Gielkens ALJ;

XX
DR WPI: 2000-106102/09.

XX
PT New avian paramyxovirus cDNA, useful for production of vaccine against
PT Newcastle disease virus

XX
PS Disclosure; Fig 3; 115pp; English.

XX
CC This invention describes a novel avian-paramyxovirus cDNA (I) which
CC comprises a nucleic acid sequence corresponding to the 5' terminal
CC end of the genome of avian-paramyxovirus allowing the generation of
CC an infectious copy of avian-paramyxovirus. The cell line is useful for
CC the production of infectious lentogenic NDV (Newcastle Disease virus)
CC without the addition of exogenous proteolytic activity. Also it is
CC possible to generate a stable transfected cell line that expresses the
CC wild-type F protein in the virus envelope therefore providing infectious
CC particles, useful in the form of a vaccine, especially against
CC respiratory and/or gastrointestinal diseases. NDV can be easily cultured
CC to very high titers in embryonated eggs. Mass culture of embryonated
CC eggs is relatively cheap. NDV vaccines are relatively stable and can be
CC simply administered by mass application methods e.g. drinking water or
CC by spraying or by aerosol formation. The natural route of infection is
CC by the respiratory and/or gastrointestinal tract which are also the major

CC routes of infection of many other poultry pathogens. NDV can induce local
 CC immunity despite the presence of circulating maternal antibody. This
 CC sequence represents a protein encoded by the NDV strain LaSota genome
 CC which is described in the method of the invention.

XX
 SQ Sequence 2204 AA;

Query Match 85.7%; Score 30; DB 21; Length 2204;
 Best Local Similarity 66.7%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKWR:6
 [: | | |]
 Db 1844 frplwr 1849

RESULT 14
 AAB53989
 ID AAB53989 standard; Protein: 90 AA.
 XX
 AC AAB53989;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human colon cancer antigen protein sequence SEQ ID NO:1529.
 XX
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; anti-infective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.

XX Homo sapiens.
 OS
 PN WO20005351-A1.
 XX
 XX 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05883.

PR 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Rosen CA, Ruben SM;

PI WPI: 2000-587534/55.

DR N-PSDB; AAC98746.

XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -

PS Claim 11; Page 2083-2084; 2104pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, anti-infective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present

CC invention.
 XX
 SQ Sequence 90 AA;

Query Match 82.9%; Score 29; DB 21; Length 90;
 Best Local Similarity 66.7%; Pred. No. 67;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKWR 6
 [: | | |]
 Db 74 fkpckw 79

RESULT 15
 AAB93531
 ID AAB93531 standard; Protein: 217 AA.

XX
 AC AAB93531;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:12886.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 12886; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dr primer and an oligonucleotide complementary

XX to the complementary strand of a polynucleotide which comprises one of

XX the 5602 nucleotide sequences defined in the specification, where the

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

xx
SQ Sequence 217 AA;

Query Match 82.9%; Score 29; DB 22; Length 217;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKPSW 5
111-1
Db 164 FKPSW 168

Search completed: February 27, 2002, 11:41:15
Job time: 454 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:17 : Search time 145.23 seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109A-7

Perfect score: 35

Sequence: 1 FKXWR 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68:*

1: piri:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	33	94.3	380	2 C64364	formate hydrogenly
2	33	94.3	412	2 E65146	hypothetical 47.4
3	33	94.3	498	2 H71279	probable fema prot
4	33	94.3	557	2 B66020	hypothetical prote
5	33	94.3	623	2 T18250	peptide transport
6	31	88.6	1035	2 S78199	probable maturase
7	30	85.7	239	2 H72667	hypothetical prote
8	30	85.7	462	2 E65131	hypothetical 47.5
9	30	85.7	462	2 B69002	probable amino aci
10	30	85.7	606	2 B69805	conserved hypothe
11	30	85.7	684	2 T31640	hypothetical prote
12	30	85.7	961	2 S67568	probable membrane
13	30	85.7	962	2 JC5808	G protein-coupled
14	30	85.7	2204	1 RRNZNV	genome polypeptin
15	29	82.9	119	2 T36433	hypothetical prote
16	29	82.9	217	2 H64956	yedG protein - Esc
17	29	82.9	222	2 F85810	hypothetical prote
18	29	82.9	301	2 T24993	hypothetical prote
19	29	82.9	395	2 H69147	LPS biosynthesis R
20	29	82.9	412	2 H75484	hypothetical prote
21	29	82.9	435	2 C86340	protein F2D10.28 [
22	29	82.9	458	2 H82504	conserved hypothe
23	29	82.9	462	2 T15052	amino acid permeas
24	29	82.9	467	2 A59268	modulation competi
25	29	82.9	473	2 G72753	hypothetical prote
26	29	82.9	477	2 S77373	hypothetical prote
27	29	82.9	488	2 B82967	probable glucose-6
28	29	82.9	522	2 T05142	glutamate--cystein
29	29	82.9	554	1 S34607	carboxylesterase (

ALIGNMENTS

RESULT 1

C64364

formate hydrogenlyase, subunit 5 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: C64364

R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
: Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999

A:Accession: C64364

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-380 <BUL>

A:Cross-references: GB:U67501; GB:L77117; NID:g2826289; PIDN:AAB98504.1; PID:g1591218

C:Genetics:

A:Map position: REV458767-457625

C:Superfamily: NADH dehydrogenase (ubiquinone) 49K protein

Query Match 94.3%; Score 33; DB 2; Length 380;
Best Local Similarity 83.3%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKXWR 6

DB 241 FKVWR 246

RESULT 2

E65146

hypothetical 47.4 kD protein in rhsB-pit intergenic region - Escherichia coli (strain

N:Alternate names: hypothetical protein f409

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999

C:Accession: E65146; S47710

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: E65146

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-412 <BLAT>

A:Cross-references: GB:AE000425; NID:U00096; PIDN:AAC76515.1; PID:g23672

A:Experimental source: Strain K-12, substrain MCL655

R:Plunkett, G.

submitted to the EMBL Data Library, March 1994

A:Reference number: S47666
A:Accession: S47710
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398, 'MLLSRCFAGSI' <PLU>
A:Cross-references: EMBL:U00039; NID:q466582; PIDN:AAB18466.1; PID:q466627
A:Note: this sequence has been corrected
C:Genetics:
A:Gene: yhiL

Query Match 94.3%; Score 33; DB 2; Length 412;
Best Local Similarity 83.3%; Pred. No. 30;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 6
III II
DB 168 FKPDWR 173

RESULT 3
H71279
probable fema protein - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: H71279
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: H71279
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-498 <COL>
A:Cross-references: GB:AE001251; GB:AE000520; NID:g3323111; PIDN:AAC65773.1; PID:g332311
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0806

Query Match 94.3%; Score 33; DB 2; Length 498;
Best Local Similarity 83.3%; Pred. No. 36;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 6
III II
DB 240 FKPKWR 245

RESULT 4
B86020
hypothetical protein 24888 [Imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: B86020
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B86020
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-557 <STO>
A:Cross-references: GB:AE005174; NID:g12518181; PIDN:AAC58622.1; GSPDB:GN00145; UWGP:248
A:Experimental source: strain O157:H7, substrain EDU933
C:Genetics:
A:Gene: 24888

Query Match 94.3%; Score 33; DB 2; Length 557;

Best Local Similarity 83.3%; Pred. No. 40;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 6
III II
DB 189 FKPDWR 194

RESULT 5
T18250
peptide transport protein - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18250
R:Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1998
A:Reference number: Z18831
A:Accession: T18250
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-623 <BAR>
A:Cross-references: EMBL:AL033503; NID:el341066; PID:el341077; PIDN:CAA22021.1
C:Genetics:
A:Gene: ptr2
A:Introns: 475/3

Query Match 94.3%; Score 33; DB 2; Length 623;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 6
III II
DB 445 FKPIWR 450

RESULT 6
S78199
probable maturase protein 3 - fission yeast (Schizosaccharomyces pombe) mitochondrion
N:Alternate names: gene cob intron 1 protein
C:Species: mitochondrion Schizosaccharomyces pombe
C:Date: 29-Jan-1998 #sequence_revision 20-Feb-1998 #text_change 21-Jul-2000
C:Accession: S78199; S10070; S67378
R:Lang, B.F.
submitted to the EMBL Data Library, August 1990
A:Description: The mitochondrial genome of the fission yeast Schizosaccharomyces pombe
chizosaccharomyces pombe and Aspergillus nidulans.
A:Reference number: S78195
A:Accession: S78199
A:Molecule type: DNA
A:Residues: 1-1035 <LAN>
A:Cross-references: EMBL:X54421
A:Experimental source: strain ad7-50h
R:Lang, B.F.; Ahne, F.; Bonen, L.
J. Mol. Biol. 184, 353-366, 1985
A:Title: The mitochondrial genome of the fission yeast Schizosaccharomyces pombe. The
e.
A:Reference number: S07274; MUID:86011547
A:Accession: S10070
A:Molecule type: DNA
A:Residues: 1-1035 <LAN>
A:Cross-references: EMBL:X02819
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC2
C:Superfamily: cytochrome b6 homology; cytochrome b homology; plastoquinol--plastocya
C:Keywords: heme; iron; metalloprotein; mitochondrion
F:1-228/Region: cytochrome b exon 1 encoded
F:10-228/Domain: cytochrome b homology #status atypical <CBH>
F:10-210/Domain: cytochrome b6 homology <CB6>
F:222-228/Domain: plastoquinol--plastocyanin reductase 17K protein homology #status a
F:222-1035/Region: cytochrome b intron encoded

Query Match 88.6%; Score 31; DB 2; Length 1035;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXWR 6
 :||| :
 Db 935 FKPSWK 940

RESULT 7
 H72667
 hypothetical protein APE0766 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: H72667
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
 A:Reference number: A72450; MUID:99310339
 A:Accession: H72667
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-239 <KAW>
 A:Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA79744.1; PID:g5104429
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0766
 C:Superfamily: Methanobacterium thermoautotrophicum conserved hypothetical protein MTH10

Query Match 85.7%; Score 30; DB 2; Length 239;
 Best Local Similarity 66.7%; Pred. No. 70;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXWR 6
 :||| :
 Db 105 FKPSWQ 110

RESULT 8
 E65131
 hypothetical 47.5 kD protein in cysG-trpS intergenic region - Escherichia coli (strain K12)
 C:Species: Escherichia coli
 C>Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000
 C:Accession: E65131
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: E65131
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-462 <BLAT>
 A:Cross-references: GB:AE000413; GB:U00096; NID:g2367215; PIDN:AAC76395.1; PID:g2367216;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: ynfM
 C:Superfamily: arginine permease

Query Match 85.7%; Score 30; DB 2; Length 462;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXWR 6
 :||| :
 Db 398 YKPLWR 403

RESULT 9
 T31640
 hypothetical protein y57A10A.r - Caenorhabditis elegans

H86002
 probable amino acid/amine transport protein yhfM [imported] - Escherichia coli (strain K12)
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: B86002
 K:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May, M.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda, Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B86002
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-462 <STO>
 A:Cross-references: GB:AE005174; NID:gl2517998; PIDN:AAG58478.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: yhfM

Query Match 85.7%; Score 30; DB 2; Length 462;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXWR 6
 :||| :
 Db 398 YKPLWR 403

RESULT 10
 B69805
 conserved hypothetical protein yfif - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: B69805
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; C.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal, I.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino, A.; Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Poll, T.M.; Portete, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Se, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchlya, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, A.; Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033
 A:Accession: B69805
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-606 <KUN>
 A:Cross-references: GB:299108; GB:AL009126; NID:g2633055; PIDN:CAB12672.1; PID:el1828
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: yfif

Query Match 85.7%; Score 30; DB 2; Length 606;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXWR 6
 :||| :
 Db 570 YKPEWR 575

RESULT 11
 T31640
 hypothetical protein y57A10A.r - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31640
R:Smyle, R.
submitted to the EMBL Data Library, September 1999
A:Reference number: 221048
A:Accession: T31640
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-684 <WIL>
A:Cross-references: EMBL:AL117195; NID:e1549729; PIDN:CAB55023.1; CESP:Y57A10A.r
A:Experimental source: clone Y57A10A
C:Genetics: ...
A:Gene: CESP:Y57A10A.r
A:Introns: 108/3; 132/3; 149/1; 180/3; 323/2; 430/1; 530/1; 584/2; 647/1

Query Match 85.7%; Score 30; DB 2; Length 684;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
||| |
DB 672 FKPKWK 677

RESULT 12
S67568
probable membrane protein YDL035C - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D2749
C:Species: Saccharomyces cerevisiae
C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 29-Oct-1999
C:Accession: S67568
R:Paulin, L.; Saren, A.M.; Laamanen, P.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67560
A:Accession: S67568
A:Molecule type: DNA
A:Residues: 1-961 <PAU>
A:Cross-references: EMBL:Z74083; NID:g1431014; PID:e252988; PID:g1431015; GSPDB:GN000004;
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YDL035C
A:Map position: 4L
C:Keywords: transmembrane protein
F:57-73/Domain: transmembrane #status predicted <TM1>
F:92-108/Domain: transmembrane #status predicted <TM2>
F:140-156/Domain: transmembrane #status predicted <TM3>
F:181-197/Domain: transmembrane #status predicted <TM4>
F:236-272/Domain: transmembrane #status predicted <TM5>
F:623-639/Domain: transmembrane #status predicted <TM6>
F:662-678/Domain: transmembrane #status predicted <TM7>

Query Match 85.7%; Score 30; DB 2; Length 961;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
||| |
DB 156 FKPKWK 161

RESULT 13
JC5808
C:protein-coupled receptor 1 - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 19-May-2000
C:Accession: JC5808
R:Yun, C.W.; Tamaki, H.; Nakayama, R.; Yamamoto, K.; Kumagai, H.
Blochem. Biophys. Res. Commun. 240, 287-292, 1997
A:Title: G-protein coupled receptor from yeast Saccharomyces cerevisiae.
A:Reference number: JC5808; MUID:98049822

A:Accession:
A:Status: nucleic acid sequence not shown
A:Molecule type: JNA
A:Residues: 1-962 <YUN>
C:Comment: This protein monitors the extracellular signal such as nutrition and trans
C:Genetics:
A:Gene: gpr1
F:56-74/Domain: transmembrane #status predicted <TM1>
F:91-109/Domain: transmembrane #status predicted <TM2>
F:139-157/Domain: transmembrane #status predicted <TM3>
F:180-198/Domain: transmembrane #status predicted <TM4>
F:255-273/Domain: transmembrane #status predicted <TM5>
F:622-640/Domain: transmembrane #status predicted <TM6>
F:661-679/Domain: transmembrane #status predicted <TM7>

Query Match 85.7%; Score 30; DB 2; Length 962;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
||| |
DB 156 FKPKWK 161

RESULT 14
JRNZNV
genome polyprotein - Newcastle disease virus (strain Beaudette C)
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: Newcastle disease virus
C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
C:Accession: A26747
R:Yusoff, K.; Millar, N.S.; Chambers, P.; Emerson, P.T.
Nucleic Acids Res. 15, 3961-3976, 1987
A:Title: Nucleotide sequence analysis of the L gene of Newcastle disease virus: homol
A:Reference number: A93665; MUID:87230982
A:Accession: A26747
A:Molecule type: mRNA
A:Residues: 1-2204 <YUS>
A:Cross-references: GB:X05399; NID:g60937; PIDN:CAA28985.1; PID:g60939
C:Genetics:
A:Gene: L
C:Superfamily: parainfluenza virus RNA-directed RNA polymerase
C:Keywords: ATP; nucleotidyltransferase

Query Match 85.7%; Score 30; DB 1; Length 2204;
Best Local Similarity 66.7%; Pred. No. 5.8e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
||| |
DB 1844 FRPLWR 1849

RESULT 15
T36433
hypothetical protein SCF43A.10c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36433
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21598
A:Accession: T36433
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-119 <SPE>
A:Cross-references: EMBL:AL096837; PIDN:CAB48897.1; GSPDB:GN00070; SCOEDB:SCF43A.10c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCF43A.10c

Thu Feb 28 09:51:11 2002

us-09-446-109a-7.rpr

Pay

Query Match 82.9%; Score 29; DB 2; Length 119;
Best Local Similarity 60.7%; Pred. No. 57;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Oy 1 FKXWR 6
Db 36 FPGWR 41

Search completed: February 27, 2002, 11:45:19
Job time: 698.56c


```

Best Local Similarity 83.3%; Pred. No. 22;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPYWR 6
Db 168 FKPMWR 173

RESULT 2
PTR2_CANAL STANDARD; PRT; 623 AA.
AC P46030;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PEPTIDE TRANSPORTER PTR2.
GN PTR2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 562 / ATCC 18804;
RX MEDLINE=95291458; PubMed=7773409;
RA Basral M.A., Lubkowitz M.A., Perry J.R., Miller D., Krainer E.,
RA Naider F.R., Becker J.M.;
RT "Cloning of a Candida albicans peptide transport gene.";
RL Microbiology 141:1147-1150(1995).
CC -1- FUNCTION: UPTAKE OF SMALL PEPTIDE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE PTR2 FAMILY OF TRANSPORTERS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U09781; AAA80167.1; -
DR InterPro; IPR000109; PTR2.
DR Pfam; PF00854; PTR2; 1.
DR PROSITE; PS01022; PTR2_1; 1.
DR PROSITE; PS01023; PTR2_2; 1.
KW Peptide transport; Transport; Transmembrane.
FT TRANSMEM 134 154 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 277 297 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 418 438 POTENTIAL.
FT TRANSMEM 448 468 POTENTIAL.
FT TRANSMEM 499 519 POTENTIAL.
FT TRANSMEM 529 549 POTENTIAL.
FT TRANSMEM 557 577 POTENTIAL.
SQ SEQUENCE 623 AA; 69941 MW; 1601FD3AE21B80EB CRC64;

Query Match 94.3%; Score 33; DB 1; Length 623;
Best Local Similarity 83.3%; Pred. No. 25;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPYWR 6
Db 445 FKPIWR 450

RESULT 3
YM91_SCHPO STANDARD; PRT; 807 AA.
ID YM91_SCHPO

Query Match 88.6%; Score 31; DB 1; Length 807;
Best Local Similarity 66.7%; Pred. No. 78;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPYWR 6
Db 707 FKPSWK 712

RESULT 4
YHFM_ECOLI STANDARD; PRT; 445 AA.
ID YHFM_ECOLI
AC P45339; P76686;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 47.5 KDA PROTEIN IN CYSG-TRPS INTERGENIC REGION.
GN YHFM OR B3370.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE ARCD/CADB/GERAB/LYSI FAMILY OF
CC PERMEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U18997; AAC58167.1; ALT SEQ.
DR EMBL: AE000413; AAC76395.1; ALT_INIT.
DR Ecogene; EG12908; ynfM.
DR InterPro: IPR002293; AA_rel_permease_1.
DR InterPro: IPR002027; Amino_acid_permease.
DR Pfam: PF00324; aa_permeases; 1.
DR Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 121 141 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT TRANSMEM 236 256 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 334 354 POTENTIAL.
FT TRANSMEM 355 375 POTENTIAL.
FT TRANSMEM 389 410 POTENTIAL.
FT TRANSMEM 417 435 POTENTIAL.
SQ SEQUENCE 445 AA: 47576 MW: 9885944.5 K1: 1; Indels 1; Mismatches 1; Gaps 0;
Query Match 85.7%; Score 30; DB 1; Length 445;
Best Local Similarity 66.7%; Pred. No. 71;
Matches 4; Conservative 1; Mismatches 1; Indels 1; Gaps 0;
QY 1 FKXWR 6
DB 381 YKPLWR 386
: : : :
: : : :
: : : :
RESULT 5
RRPL_NDVB STANDARD; PRT: 2204 AA.
ID RRPL_NDVB
AC P11205;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE (L PROTEIN).
GN L.
OS Newcastle disease virus (strain Beaudette C/45) (NDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
OX NCBI_TaxID=11178;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=87230982; PubMed=3035486;
RA Yusoff K., Millar N.S., Chambers P., Emmerson P.T.;
RT "Nucleotide sequence analysis of the L gene of Newcastle disease
RT virus: homologues with Sendai and vesicular stomatitis viruses.";

RL Nucleic Acids Res. 15:3961-3976(1987).
CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P
CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC -1- SIMILARITY: WITH L PK-PROTEIN OF OTHER PARAMYXOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X05399; CAA28985.1;
DR PIR: A26747; RBNZNV.
DR InterPro: IPR001016; Paramyx_RNA_pol.
DR Pfam: PF00946; Paramyx_RNA_pol; 1.
KW Transferase; RNA-directed RNA polymerase.
SQ SEQUENCE 2204 AA: 248822 MW: 6578674.9 D04802C CRC64;
Query Match 85.7%; Score 30; DB 1; Length 2204;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKXWR 6
DB 1844 FRPLWR 1849
: : : :
: : : :
: : : :
RESULT 6
YEDK_ECOLI STANDARD; PRT: 217 AA.
ID YEDK_ECOLI
AC P76318; 1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 24.5 KDA PROTEIN IN AMYA-FLIE INTERGENIC REGION.
GN YEDK OR B1931.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=JALL.
RX MEDLINE=93381452; PubMed=8371104;
RA Raha M., Kihara M., Kawagishi I., Macnab R.M.;
RA "Organization of the Escherichia coli and Salmonella typhimurium
RA chromosomes between flagellar regions IIIa and IIb, including a
RA large non-coding region."
RT J. Gen. Microbiol. 139:1401-1407(1993).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- SIMILARITY: TO YEAST YMR114C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```
CC EMBL; L13279; NOT_ANNOTATED_CDS.
DR EMBL; AE000285; AAC74998.1; -.
DR EcoGene; EG13278; yedK.
DR InterPro; IPR003738; DUF159.
DR Pfam; PF02586; DUF159; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 213 217 TRASN-> NOGAELIOPV (IN REF. 1).
SQ SEQUENCE 217 AA; 24500 MW; 4F26C95DB3802900 CRC64;

Query Match: 82.9%; Score 29; DB 1; Length 217;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKXWR 6
DB 87 FKPLWQ 92

RESULT 7
SPS2_MOUSE STANDARD; PRT; 452 AA.
AC P97364;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SELENIDE, WATER DIKINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)
DE (SELENIUM DONOR PROTEIN 2).
GN SPS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=96017645; PubMed=7588067;
RA Guimaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,
RA Lee F., McClanahan T.;
RT "A new approach to the study of haematopoietic development in the
RT yolk sac and embryoid bodies."
RL Development 121:3335-3346(1995).
RN [2]
RX MEDLINE=97140286; PubMed=8986768;
RA Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,
RA Zlotnik A.;
RT "Identification of a novel self homolog from eukaryotes, bacteria,
RT and archaea: is there an autoregulatory mechanism in selenocysteine
RT metabolism?";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).
CC -1- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.
CC -1- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O -> AMP + SELENOPHOSPHATE
CC + PHOSPHATE
CC -1- COPACITOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
CC ENCODED BY THE OPAL CODON, UGA.
CC -1- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.
CC CLASS 1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U43285; AAC53024.1; -.
CC MGD; MGI:108388; Sps2.
CC InterPro; IPR000728; AIRS_related.
CC Pfam; PF00586; AIRS; 1.
KW Transferase; Selenium; Selenocysteine; ATP-binding.

```

```
FT ACT_SITE 63 63 POTENTIAL.
FT SE_CYS 63 63
FT SITE 66 66 IMPORTANT FOR CATALYTIC ACTIVITY (BY
FT SIMILARITY).
FT NP_BIND 322 328 ATP (POTENTIAL).
FT DOMAIN 2 9 POLY-ALA.
FT DOMAIN 433 440 POLY-ALA.
SQ SEQUENCE 452 AA; 47786 MW; 9DA6F7250CFE80B4 CRC64;

Query Match: 82.9%; Score 29; DB 1; Length 452;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKXWR 6
DB 48 FSPSWR 53

RESULT 8
Y264_SYNY3 STANDARD; PRT; 477 AA.
AC P73436;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 54.0 KDA PROTEIN SLL1464.
GN SLL1464.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosokuchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions."
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90906; BAA17476.1; -.
CC InterPro; IPR003846; UPF0061.
CC Pfam; PF02696; UPF0061; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 477 AA; 54041 MW; 81F6899B1A6D613C CRC64;

Query Match: 82.9%; Score 29; DB 1; Length 477;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKXWR 6
DB 388 FSPSWR 393

RESULT 9
CSHL_BRAJU STANDARD; PRT; 514 AA.
ID GSH1_BRAJU
AC O23736; Q43389;

```

DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GLUTAMATE--CYSTEINE LIGASE, CHLOROPLAST PRECURSOR (EC 6.3.2.2) (GAMMA-GLUTAMYL-CYSTEINE SYNTHETASE) (GAMMA-ECS) (GCS).
 GN GSH1.
 DE Brassica juncea (Leaf mustard) (Indian mustard).
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 NCBI_TaxID=3707;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP TISSUE=ROOT;
 RC MEDLINE=98281577; PubMed=9620267;
 RX Schaefer H.J., Haag-Kerwer A., Rausch T.;
 RA "cDNA cloning and expression analysis of genes encoding GSH synthesis
 RT in roots of the heavy-metal accumulator Brassica juncea L.: evidence
 RT for Cd-induction of a putative mitochondrial gamma-glutamylcysteine
 RT synthetase isoform.";
 RL Plant Mol. Biol. 37:87-97(1998).
 RN [2]
 RN SEQUENCE OF 126-342 FROM N.A.
 RP STRAIN=CV. VITTASSO; TISSUE=Leaf;
 RX MEDLINE=97227950; PubMed=9119067;
 RA Schaefer H.J., Greiner S., Rausch T., Haag-Kerwer A.;
 RT "In seedlings of the heavy metal accumulator Brassica juncea, Cu2+
 RT differentially affects transcript amounts for gamma-glutamylcysteine
 RT synthetase (gamma-ECS) and metallothionein (MT2).";
 RL FEBS Lett. 404:216-220(1997).
 CC -|- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + L-CYSTEINE = ADP +
 CC ORTHOPHOSPHATE + GAMMA-L-GLUTAMYL-L-CYSTEINE.
 CC -|- PATHWAY: FIRST AND RATE-LIMITING STEP IN DE NOVO GLUTATHIONE
 CC BIOSYNTHESIS.
 CC -|- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).
 CC -|- MISCELLANEOUS: IN PLANTS, THERE ARE TWO FORMS OF THIS ENZYME, ONE
 CC CYTOSOLIC AND THE OTHER CHLOROPLASTIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See: <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y10848; CAA71801.1;
 DR EMBL: X95563; CAA64808.1;
 KW Glutathione biosynthesis; Ligase; Chloroplast; Transit peptide.
 FT TRANSIT 1 55 CHLOROPLAST (POTENTIAL).
 FT CHAIN 56 514 GLUTAMATE--CYSTEINE LIGASE.
 FT CONFLICT 139 139 K -> R (IN REF. 2).
 FT CONFLICT 215 215 T -> I (IN REF. 2).
 FT CONFLICT 257 257 M -> T (IN REF. 2).
 SQ SEQUENCE 514 AA: 57903 MW: 07C71CB13E785FA8 CRC64;
 Query Match 82.9%; Score 29; DB 1; Length 514;
 Best Local Similarity 66.7%; Pred No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FRPKWR 6
 I: I I I
 Db 204 FQPKWR 209
 RESULT 10
 ID GSH1ARATH STANDARD; PRT: 522 AA.
 AC P46309: O82759; P92951;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GLUTAMATE--CYSTEINE LIGASE, CHLOROPLAST PRECURSOR (EC 6.3.2.2) (GAMMA-GLUTAMYL-CYSTEINE SYNTHETASE) (GAMMA-ECS) (GCS).
 GN GSH1 OR AT4G231100 OR F7H19.290.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NCBI_TaxID=3702;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. COLUMBIA; TISSUE=Leaf;
 RX MEDLINE=95023984; PubMed=7937837;
 RA May M.J., Leaver C.J.;
 RT "Arabidopsis thaliana gamma-glutamylcysteine synthetase is
 RT structurally unrelated to mammalian, yeast, and Escherichia coli
 RT homologs.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:10059-10063(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. LANDSBERG ERECTA;
 RA Ullmann P., Condet L., Bach T.J.;
 RT "Isolation of an Arabidopsis thaliana cDNA encoding a putative
 RT gamma-glutamylcysteine synthetase by complementation of a GSHI
 RT deficient yeast mutant-glutamylcysteine synthetase.";
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. COLUMBIA;
 RA Cobbett C.S., May M.J., Howden R., Rolls B.;
 RT "The glutathione-deficient, cadmium-sensitive mutant, cad2-1, of
 RT Arabidopsis thaliana is deficient in gamma-glutamylcysteine
 RT synthetase.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terry N.,
 RA Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller T.,
 RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelie D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohnel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenegger T., Bothe G., Ransperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buysschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
 RA Pettett A., Rajandream M.-A., Lyne M., Benes V., Reclmann S.,
 RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loelmer T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Grandrath K., Dauner D., Herzl A.,
 RA Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chedford F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacón D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Hannon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,

RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., All J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 thaliana.";
 RL Nature 402:769-777(1999).
 CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + L-CYSTEINE -> ADP +
 CC ORTHOPHOSPHATE + GAMMA-L-GLUTAMYL-L-CYSTEINE.
 CC -!- ENZYME REGULATION: FEEDBACK INHIBITION BY GLUTATHIONE.
 CC -!- PATHWAY: FIRST AND RATE-LIMITING STEP IN DE NOVO GLUTATHIONE
 CC BIOSYNTHESIS.
 CC -!- SUBUNIT: MONOMER (PROBABLE).
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST (PROBABLE).
 CC -!- TISSUE SPECIFICITY: ABUNDANT IN LEAVES AND ROOTS.
 CC -!- MISCELLANEOUS: IN PLANTS, THERE ARE TWO FORMS OF THIS ENZYME, ONE
 CC CYTOSOLIC AND THE OTHER CHLOROPLASTIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 229490; CAAB2626.1; -;
 DR EMBL: Y09944; CAA71075.1; -;
 DR EMBL: AF088299; AAD14544.1; -;
 DR EMBL: AL031018; CAA19826.1; -;
 DR EMBL: AL161558; CAB79265.1; -;
 DR Mendel: 17259; Arab: 2621:17259.
 DR Mendel: 33329; Arab: 2621:33329.
 DR Mendel: 36232; Arab: 2621:36232.
 KW Glutathione biosynthesis; Ligase; Chloroplast; Transit peptide.
 FT TRANSIT 1 ?
 FT CHAIN 1 ?
 FT VARIANT 22 22
 FT VARIANT 48 48
 FT CONFLICT 491 522
 FT YACGALGVQVMKTKRRSRVRAAALVRKMDVNRKCL
 FT (IN REF. 1).
 FT SQ SEQUENCE 522 AA; 58562 MW; CCBF13C6F44E0EF7 CRC64;
 Query Match 82.9%; Score 29; DB 1; Length 522;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 FKPKWR 6
 Db 212 FQPKWR 217
 RESULT 11
 ESTM_MOUSE
 ID ESTM_MOUSE STANDARD; PRT; 554 AA.
 AC Q63880;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE LIVER CARBOXYLESTERASE PRECURSOR (EC 3.1.1.1) (ES-MALE) (ESTERASE-31).
 GN ES31.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-BALB/CJ X DBA/2J; TISSUE=Liver;

RX MEDLINE-93:24538; PubMed=7916639;
 RA Aida K., Moore R., Negishi M.;
 RT "Cloning and nucleotide sequence of a novel, male-predominant
 RT carboxylesterase in mouse liver.";
 RL Biochim. Biophys. Acta 1174:72-74(1993).
 CC -!- FUNCTION: INVOLVED IN THE DETOXIFICATION OF XENOBIOTICS AND IN
 CC THE ACTIVATION OF ESTER AND AMIDE PRODRUGS.
 CC -!- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O -> AN ALCOHOL
 CC + A CARBOXYLIC ANION.
 CC -!- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE, LUMEN OF ENDOPLASMIC
 CC RETICULUM.
 CC -!- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S64130; AAB27606.1; -;
 DR HSPD: P21836; IMAH.
 DR MGD: MGI:102773; Es31
 DR InterPro: IPR002018; Carboxylesterase_B.
 DR InterPro: IPR000379; Est_lip_thioest_actsite.
 DR Pfam: PF001135; Coesterase; 1.
 DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
 KW Glycoprotein; Hydrolase; Serine esterase; Endoplasmic reticulum;
 KW Signal; Multigene family.
 FT SIGNAL 1 14
 FT CHAIN 15 554
 FT ACT_SITE 215 215
 FT ACT_SITE 443 443
 FT DISULFID 83 110
 FT DISULFID 267 278
 FT SITE 551 554
 FT CARBOHYD 294 294
 FT CARBOHYD 300 300
 FT CARBOHYD 381 381
 FT SEQUENCE 554 AA; 61509 MW; 028D89806F3CAAD CRC64;
 Query Match 82.9%; Score 29; DB 1; Length 554;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 FKPKW 5
 Db 434 FKPAW 438
 RESULT 12
 HSPD_BRAJA
 ID HSPD_BRAJA STANDARD; PRT; 151 AA.
 AC O69241;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPD.
 GN HSPD.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Narberhaus F., Weighofer W., Fischer H.M., Hennecke H.;
 RT "Identification of the Bradyrhizobium japonicum degp gene as part of
 RT an operon containing small heat shock protein genes.";
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)

CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ003064; CAA05835.1;
 DR InterPro: IPR002068; Crystallin_HSP20.
 DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SQ SEQUENCE 151 AA; 17272 MW; FB44EF94FB599EB4 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 64;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 FKPKXR 6
 DB 6 FSPLMR 11
 ID HSPH_BRAJA STANDARD; PRT; 151 AA.
 AC O86110;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DE SMALL HEAT SHOCK PROTEIN HSPH.
 GN HSPH.
 OS Bradyrhizobium japonicum.
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Muenchbach M., Nocker A., Narberhaus F.;
 RT "Occurrence of a superfamily of small heat shock proteins in
 RT Bradyrhizobium japonicum and other Rhizobium species: a plant-like
 RT phenomenon";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ010144; CAA09014.1;
 DR InterPro: IPR002068; Crystallin_HSP20.
 DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SQ SEQUENCE 151 AA; 17098 MW; DFE9809310AF6A28 CRC64;
 Query Match 80.0%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 64;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 FKPKXR 6
 DB 6 FSPLMR 11

RESULT 14
 YLF6_CAEEL STANDARD; PRT; 162 AA.
 ID YLF6_CAEEL
 AC Q03598;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 18.5 KDA PROTEIN C40H1.6 IN CHROMOSOME III.
 GN C40H1.6.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 KC MEDLINE=94150718; PubMed=7906398;
 KA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 KA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 KA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 KA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 KA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 KA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 KA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 KA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
 KA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 KA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 KA Wohlman P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans";
 RT Nature 368:32-38(1994).
 CC -!- SIMILARITY: STRONG, TO HUMAN CGI-126.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z19154; CAA79557.1;
 DR PIR: S28301; S28301.
 DR WormPep: C40H1.6; CE00114.
 KW Hypothetical protein.
 SQ SEQUENCE 162 AA; 18537 MW; 8DC03CE3BCE79D55 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 162;
 Best Local Similarity 80.0%; Pred. No. 68;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKPKXR 5
 DB 120 FKPLW 124

RESULT 15
 YLF6_HUMAN STANDARD; PRT; 167 AA.
 ID YLF6_HUMAN
 AC Q9V3C8;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN CGI-126.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-20272150; PubMed-10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RL Caenorhabditis elegans by comparative proteomics.";
CC Genome Res. 10:703-713(2000).
CC -1- SIMILARITY: STRONG, TO C.ELEGANS C40H1.6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF151884; AAD34121.1; -.
KW Hypothetical protein.
SQ SEQUENCE 167 AA; 19458 MW; 1675D9187DC43E14 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 167;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;
Oy 1 FKPKW 5
Db 121 FKPLW 125

Search completed: February 27, 2002, 11:42:45
Job time: 544 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:07 : Search time 281.76 Seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109a-7
Perfect score: 35
Sequence: 1 FKXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertibrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	97.1	250	Q9EUM2	Q9EUM2 corynebacte
2	33	94.3	380	Q57935	Q57935 methanococc
3	33	94.3	498	Q83784	Q83784 treponema p
4	33	94.3	623	Q9URL7	Q9URL7 candida alb
5	33	94.3	917	Q9XP8	Q9XP8 drosophila
6	33	94.3	917	Q9VY94	Q9VY94 drosophila
7	31	88.6	221	Q9XJ12	Q9XJ12 oryza sativ
8	30	85.7	204	Q9DJF2	Q9DJF2 mus musculu
9	30	85.7	239	Q9YE03	Q9YE03 aeropyrum p
10	30	85.7	440	Q9D9R1	Q9D9R1 mus musculu
11	30	85.7	606	Q31566	Q31566 bacillus su
12	30	85.7	610	Q52961	Q52961 bacillus su
13	30	85.7	684	Q9NA80	Q9NA80 caenorhabdi
14	30	85.7	961	Q12361	Q12361 saccharomyc
15	30	85.7	1232	Q9LJQ1	Q9LJQ1 arabidopsis
16	30	85.7	2204	Q90341	Q90341 newcastle d
17	30	85.7	2204	Q9WMH6	Q9WMH6 newcastle d
18	30	85.7	2204	Q9DLD3	Q9DLD3 newcastle d
19	30	85.7	2454	Q9UVP2	Q9UVP2 emericella

20	30	85.7	2454	3	Q9UVP56	Q9UVP56 emericella
21	29	82.9	119	2	Q9XAA5	Q9XAA5 streptomyce
22	29	82.9	138	2	Q9ZEW4	Q9ZEW4 klebsiella
23	29	82.9	143	5	Q9G402	Q9G402 plasmidum
24	29	82.9	162	2	Q9GJP0	Q9GJP0 pasteurella
25	29	82.9	164	5	Q9NJI1	Q9NJI1 giardia lam
26	29	82.9	224	5	Q9NZY6	Q9NZY6 caenorhabdi
27	29	82.9	233	2	Q9F2R9	Q9F2R9 streptomyce
28	29	82.9	244	10	Q9FS78	Q9FS78 triticum ae
29	29	82.9	262	5	Q9NND2	Q9NND2 leishmania
30	29	82.9	296	2	Q33353	Q33353 mycobacteri
31	29	82.9	301	5	Q22568	Q22568 caenorhabdi
32	29	82.9	395	1	Q26470	Q26470 methanobact
33	29	82.9	398	2	Q9A710	Q9A710 caulobacter
34	29	82.9	412	2	Q9RWE3	Q9RWE3 deinococcus
35	29	82.9	435	10	Q9LM75	Q9LM75 arabidopsis
36	29	82.9	436	10	Q82665	Q82665 brassica ju
37	29	82.9	445	3	Q9C1B2	Q9C1B2 gibberella
38	29	82.9	447	3	Q9A088	Q9A088 fusarium sp
39	29	82.9	447	3	Q9C1B9	Q9C1B9 fusarium sp
40	29	82.9	458	2	Q9KN89	Q9KN89 vibrio chol
41	29	82.9	462	10	Q40414	Q40414 nicotiana s
42	29	82.9	467	2	Q52729	Q52729 rhizobium e
43	29	82.9	473	1	Q9YGB1	Q9YGB1 aeropyrum p
44	29	82.9	488	2	Q9H7C7	Q9H7C7 pseudomonas
45	29	82.9	557	3	Q42902	Q42902 schizosacch

ALIGNMENTS

RESULT 1
Q9EUM2 ID Q9EUM2 PRELIMINARY: PRT; 250 AA.
AC Q9EUM2;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE YCG4L.
GN YCG4L.
GS Corynebacterium glutamicum (Brevibacterium flavum).
GC Plasmid R-plasmid pCG4.
GC Bacteria: Firmicutes; Actinobacteria: Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID: 1718;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN:ATCC31830;
KA Tauch A., Mueller A., Kalinowski J.;
KT "DNA sequ and genetic organization of the integron-carrying R-
PT plasmid p of Corynebacterium glutamicum.";
RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF111555; AAC00294.1;
DR InterPro:IPR002145; CopG_HTH_4.
DR Plasm: PF01442; HTH_4; 1.
KW Plasmid.
SQ SEQUENCE 250 AA; 28804 MW; 195C86367B483157 CRC64;

Query Match 97.1%; Score 34; DB 2; Length 250;
Best Local Similarity 83.3%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKXWR 6
DB 214 FKXWR 219

RESULT 2
ID Q57935 PRELIMINARY: PRT; 380 AA.
AC Q57935;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PUTATIVE FORMATE HYDROGENLYASE SUBUNIT 5 (PFL SUBUNIT 5) (HYDROGENASE-
 GN 3 COMPONENT E).
 GN MJ0515.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_TaxID=21190;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervilange A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii*;
 RL Science 273:1058-1073(1996).
 CC -1- COFACTOR: NICKEL (BY SIMILARITY).
 CC -1- PATHWAY: HYDROGEN METABOLISM; PFL PATHWAY (BY SIMILARITY).
 CC -1- SUBUNIT: PFL COMPRISES OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
 CC ELECTRON CARRIERS AND A HYDROGENASE (ISOENZYME 3). IN THIS NON-
 CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBODIOXIDE FROM
 CC FORMATE ARE RELEASED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 49 KDA SUBUNIT FAMILY.
 DR EMBL: U67501; AAB9504.1; -;
 DR TIGR: MJ0515; -;
 DR InterPro: IPR001135; Complex1_49kd.
 DR InterPro: IPR001501; NiFeSe_Hases.
 DR Pfam: PF00346; complex1_49kd; 1.
 DR Pfam: PF00374; NiFeSe_Hases; 1.
 DR PROSITE: PS00535; COMPLEX1_49K; UNKNOWN_1.
 DR PROSITE: PS00507; NI_HGENASE_L1; UNKNOWN_1.
 DR Hypothetical protein; Oxidoreductase; NAD; Iron-sulfur; 4Fe-4S;
 KW Nickel; Complete proteome.
 SQ SEQUENCE 380 AA; 43638 MW; 3EAD16D67834A76F CRC64;

Query Match 94.3%; Score 33; DB 1; Length 380;
 Best Local Similarity 83.3%; Pred. No. 81;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 241 FKPVNR 246

RESULT 3
 ID 083784 PRELIMINARY; PRT; 498 AA.
 AC 083784;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE FENX PROTEIN, PUTATIVE.
 DE TP0806.
 GN Treponema pallidum.
 OS Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OC NCBI_TaxID=160;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=NICHOLS;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khalak H., D., Howell J.K., Chidamb.
 RA McDonald, P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatcher B., Hurst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT *Complete genome sequence of Treponema pallidum, the syphilis
 spirochete*;
 XL Science 281:375-388(1998).
 DR EMBL: AE001251; AAC65773.1; -;
 DR TIGR: TP0806; -;
 DR InterPro: IPR003447; FemAB.
 DR Pfam: PF02388; FemAB; 1.
 DR Complete proteome.
 KW SEQUENCE 498 AA; 56074 MW; 81710E41264431D8 CRC64;

Query Match 94.3%; Score 33; DB 2; Length 498;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 240 FKPKWR 245

RESULT 4
 ID Q9URL7 PRELIMINARY; PRT; 623 AA.
 AC Q9URL7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PEPTIDE TRANSPORT PROTEIN.
 GN PTR2.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=1161;
 RA Murphy L., Harris D.;
 RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RA Barrell B.G., Rajandream M.A.;
 RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=1161;
 EX MEDLINE=97435544; PubMed=9290243;
 RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
 RT "A Candida albicans Genome Project: Cosmid Contigs, Physical Mapping,
 and Gene Isolation";
 XL Fungal Genet. Biol. 21:308-314(1997).
 DR EMBL: AL033503; CAA22021.1; -;
 DR InterPro: IPR000109; PTR2.
 DR Pfam: PF00854; PTR2; 1.
 DR PROSITE: PS01022; PTR2_1; 1.
 DR PROSITE: PS01023; PTR2_2; UNKNOWN_1.
 SQ SEQUENCE 623 AA; 69943 MW; 8543A3B1F7E7363E CRC64;

Query Match 94.3%; Score 33; DB 3; Length 623;
 Best Local Similarity 83.3%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 445 FKPIWR 450

RESULT 5

```

Q9XP8
ID Q9XP8 PRELIMINARY; PRT: 917 AA.
AC Q9XP8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GAMMA-TUBULIN RING PROTEIN DGRIP91.
GN L(1)DD4 OR DGRIP91 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99156983; PubMed=10037793;
RA Oegema K., Wiese C., Martin O., Milligan R.A., Iwamatsu A.,
RA Mitchison T.J., Zheng Y.;
RT "Characterization of two related Drosophila gamma-tubulin complexes
RT that differ in their ability to nucleate microtubules.";
RL J. Cell Biol. 144:721-733(1999).
DR EMBL; AF118380; AAD27817.1; -.
DR FlyBase; FBgn0001612; l(1)dd4.
DR InterPro; IPR000634; dehydrtse_ser_thr.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR: 1.
SQ SEQUENCE 917 AA; 103819 MW; EBF2F41E35049D27 CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
DB 657 FKPLWR 662

RESULT 6
ID Q9VY94 PRELIMINARY; PRT: 917 AA.
AC Q9VY94;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GRIP91 PROTEIN.
GN L(1)DD4 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

```

```

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strouy R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003493; AAF48309.1; -.
DR FlyBase; FBgn0001612; l(1)dd4.
DR InterPro; IPR000634; dehydrtse_ser_thr.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR: 1.
SQ SEQUENCE 917 AA; 103706 MW; 6AEE88C211D256BB CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
DB 657 FKPLWR 662

RESULT 7
ID Q9XJ12 PRELIMINARY; PRT: 221 AA.
AC Q9XJ12;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. NIPPONBARE;
RA Sasaki T., Nagamura Y., Yamamoto K.;
RT "Oryza sativa nipponbare(CA3) genomic DNA, chromosome 6, PAC
RT clone:P0680A03.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB023452; BAA78741.1; -.
SQ SEQUENCE 221 AA; 24502 MW; DC24FF370660BC15 CRC64;

Query Match 88.6%; Score 31; DB 10; Length 221;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
DB 203 FAPTWR 208

RESULT 8
ID Q9D3F2 PRELIMINARY; PRT: 204 AA.
AC Q9D3F2;

```

```

DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 5830420C20RIK PROTEIN.
GN 5830420C20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=THYMUS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK017942; BAB31014.1; -.
DR MGD: MGI:1923275; 5830420C20RIK.
SQ SEQUENCE 204 AA; 22892 MW; 98AF03DF7FC7B066 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 204;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
Db 29 FRPRWR 34

RESULT 9
ID Q9YE03 PRELIMINARY; PRT: 239 AA.
AC Q9YE03;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 24.8 KDA PROTEIN APE0766.
GN APE0766.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KJ;
RX MEDLINE=99310339; PubMed=10382966;
RA Kwarabavasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankaï A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).

```

```

DR EMBL: AP000060; BAA79744.1; -.
DR InterPro: IPR001454; Hydrolase.
DR InterPro: IPR000150; Hypothes. cof.
DR Pfam: PF00702; Hydrolase; 1.
DR PROSITE: PS01229; COF_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 239 AA; 24758 MW; EA30A3C0FD86B3DE CRC64;

Query Match 85.7%; Score 30; DB 11; Length 239;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
Db 105 FKPSWQ 110

RESULT 10
ID Q9D9R1 PRELIMINARY; PRT: 440 AA.
AC Q9D9R1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:1700030N20, FULL INSERT SEQUENCE.
GN 5830420C20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK006565; BAB24655.1; -.
DR MGD: MGI:1923275; 5830420C20RIK.
SQ SEQUENCE 440 AA; 49566 MW; 5B5DA87E19879644 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 440;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
Db 39 FRPRWR 44

RESULT 11
ID Q31566 PRELIMINARY; PRT: 606 AA.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:07 ; Search time 281.76 Seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109a-7
Perfect score: 35
Sequence: 1 FKPMWR 6

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL17.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	97.1	250	2 09EUM2	09eum2 corynebacte
2	33	94.3	380	1 057935	057935 methanococc
3	33	94.3	498	2 083784	083784 treponema p
4	33	94.3	623	3 09UKL7	09ur17 candida alb
5	33	94.3	917	5 09XYP8	09xyp8 drosophila
6	33	94.3	917	5 09VY94	09vy94 drosophila
7	31	88.6	221	10 09XJ12	09xj12 oryza sativ
8	30	85.7	204	11 09D3F2	09d3f2 mus musculu
9	30	85.7	239	1 09YK03	09ye03 acropyrum p
10	30	85.7	440	11 09D9R1	09d9r1 mus musculu
11	30	85.7	606	2 031566	031566 bacillus su
12	30	85.7	610	2 052961	052961 bacillus su
13	30	85.7	684	5 09NA80	09na80 caenorhabdi
14	30	85.7	961	3 012361	012361 saccharomyc
15	30	85.7	1232	10 091JQ1	091jq1 arabidopsis
16	30	85.7	2204	12 090341	090341 newcastle d
17	30	85.7	2204	12 09MMH6	09mmh6 newcastle d
18	30	85.7	2204	12 09DLD3	09dld3 newcastle d
19	30	85.7	2454	3 09UVP2	09uvp2 emericella

20	30	85.7	2454	3 09UVP56	09uvp56 emericella
21	29	82.9	119	2 09XAA5	09xaa5 streptomyce
22	29	82.9	138	2 09ZEW4	09zew4 klebsiella
23	29	82.9	143	5 096402	096402 plasmodium
24	29	82.9	162	2 09CJP0	09cjp0 pasteurella
25	29	82.9	164	5 09NJY1	09njy1 giardia lam
26	29	82.9	224	5 09N2Y6	09n2y6 caenorhabdi
27	29	82.9	233	2 09F2R9	09f2r9 streptomyce
28	29	82.9	244	10 09FS78	09fs78 triticum ac
29	29	82.9	262	5 09NND2	09nnd2 leishmania
30	29	82.9	296	2 033353	033353 mycobacteri
31	29	82.9	301	5 022568	022568 caenorhabdi
32	29	82.9	395	1 026470	026470 methanobact
33	29	82.9	398	2 09A710	09a710 caulobacter
34	29	82.9	412	2 09RWE3	09rwe3 deinococcus
35	29	82.9	435	10 09LM75	09lm75 arabidopsis
36	29	82.9	436	10 082665	082665 brassica ju
37	29	82.9	445	3 09C1B2	09c1b2 qibberella
38	29	82.9	447	3 094088	094088 fusarium sp
39	29	82.9	447	3 09C1B9	09c1b9 fusarium sp
40	29	82.9	458	2 09KN89	09kn89 vibrio chol
41	29	82.9	462	10 040414	040414 nicotiana s
42	29	82.9	467	2 052729	052729 rhizobium e
43	29	82.9	473	1 09YGB1	09ygb1 aeropyrum p
44	29	82.9	488	2 09HTC7	09htc7 pseudomonas
45	29	82.9	557	3 042902	042902 schizosacch

ALIGNMENTS

RESULT 1
Q9EUM2 ID Q9EUM2 PRELIMINARY: PRT: 250 AA.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE YCG4L.
GN YCG4L.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OG Plasmid R-plasmid pCG4.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCRI_TaxID=1718;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC11830;
RA Tauch A., Puhler A., Kallinowski J.;
RT "DNA sequences and genetic organization of the integron-carrying R-
plasmid pCG4 of Corynebacterium glutamicum.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF164956; AAG00294.1;
DR InterPro: IPR002145; CopG_HTH_4.
DR 71am: PF01402; HTH_4; 1.
KW plasmid.
SQ SEQUENCE 250 AA: 28804 MW: 195C86367B483157 CRC64;

Query Match 97.1%; Score 34; DB 2; Length 250;
Best Local Similarity 83.3%; Pred. No. 35;
Matches: 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMWR 5
111 11
DB 214 FKPMWR 219

RESULT 2
Q9EUM2 ID Q9EUM2 PRELIMINARY: PRT: 380 AA.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE YCG4L.
GN YCG4L.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OG Plasmid R-plasmid pCG4.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCRI_TaxID=1718;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC11830;
RA Tauch A., Puhler A., Kallinowski J.;
RT "DNA sequences and genetic organization of the integron-carrying R-
plasmid pCG4 of Corynebacterium glutamicum.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF164956; AAG00294.1;
DR InterPro: IPR002145; CopG_HTH_4.
DR 71am: PF01402; HTH_4; 1.
KW plasmid.
SQ SEQUENCE 250 AA: 28804 MW: 195C86367B483157 CRC64;

01-JAN-1998 (TrEMBLrel. 05, Created)
 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 PUTATIVE FORMATE HYDROGENLYASE SUBUNIT 5 (FHL SUBUNIT 5) (HYDROGENASE-3 COMPONENT E).
 DE
 GN M0515.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota: Methanococcales: Methanococcaceae;
 OX Methanococcus.
 QX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 KC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Karlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Gocayne J.D.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glouck A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhlmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.;
 RL Science 273:1058-1073(1996).
 CC -1- COFACTOR: NICKEL (BY SIMILARITY).
 CC -1- PATHWAY: HYDROGEN METABOLISM: FHL PATHWAY (BY SIMILARITY).
 CC -1- SUBUNIT: FHL COMPRISES OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
 CC ELECTRON CARRIERS AND A HYDROGENASE (ISOENZYME 3). IN THIS NON-
 CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBODIOXIDE FROM
 CC FORMATE ARE RELEASED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 49 KDA SUBUNIT FAMILY.
 DR EMBL: U67501; AAB98504.1; -;
 DR TIGR: M0515; -;
 DR InterPro: IPR001135; Complex1_49kd.
 DR InterPro: IPR001501; NiFeSe_Hases.
 DR Pfam: PF00346; complex1_49kd; 1.
 DR Pfam: PF00374; NiFeSe_Hases; 1.
 DR PROSITE: PS00535; COMPLEX1_49K; UNKNOWN_1.
 DR PROSITE: PS00507; NI_GENASE_L1; UNKNOWN_1.
 KW Hypothetical protein: Oxidoreductase; NAD; Iron-sulfur; 4Fe-4S;
 KW Nickel; Complete proteome.
 SQ SEQUENCE 380 AA; 43638 MW; 3EAD16D67834A76F CRC64;

Query Match 94.3%; Score 33; DB 1; Length 380;
 Best Local Similarity 83.3%; Pred. NO. 81;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKWR 6
 III II
 DB 241 FKPVWR 246

RESULT 3
 ID 083784 PRELIMINARY; PRT: 498 AA.
 AC 083784;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE FHLA PROTEIN, PUTATIVE.
 GN TP0806.
 OS Treponema pallidum.
 OC Bacteria: Spirochaetales: Spirochaetaceae; Treponema.
 OX NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NICHOLS;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Soderjgren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artisch P., Rowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT *Complete genome sequence of Treponema pallidum, the syphilis
 agent "Spirochete";
 RL Science 281:375-388(1998).
 DR EMBL: A6001251; AAC65773.1; -;
 DR TIGR: TP0806; -;
 DR InterPro: IPR003447; FemAB.
 DR Pfam: PF02388; FemAB; 1.
 KW Complete proteome.
 SQ SEQUENCE 498 AA; 56074 MW; 81710E41264431H8 CRC64;
 Query Match 94.3%; Score 33; DB 2; Length 498;
 Best Local Similarity 83.3%; Pred. NO. 1e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKWR 6
 III II
 DB 240 FKPVWR 245

RESULT 4
 ID 09URL7 PRELIMINARY; PRT: 623 AA.
 AC 09URL7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PEPTIDE TRANSPORT PROTEIN.
 GN PTR2
 OS Candida albicans (Yeast).
 OC Eukaryota: Fungi; Ascomycota; Saccharomycotina: Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RA Murphy L., Harris D.;
 RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RA Barrall B.G., Rajandream M.A.;
 RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1161;
 RX MEDLINE=97435544; PubMed=9290243;
 RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;
 RT *A Candida albicans Genome Project: Cosmid Contigs, Physical Mapping,
 RT and Gene Isolation.;
 RL Fungal Genet. Biol. 21:308-314(1997).
 DR EMBL: AL033503; CAA22021.1; -;
 DR InterPro: IPR00109; PTR2.
 DR Pfam: PF00854; PTR2; 1.
 DR PROSITE: PS01022; PTR2_1; 1.
 DR PROSITE: PS01023; PTR2_2; UNKNOWN_1.
 SQ SEQUENCE 623 AA; 69943 MW; 8543A381F7E7363E CRC64;

Query Match 94.3%; Score 33; DB 3; Length 623;
 Best Local Similarity 83.3%; Pred. NO. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKWR 6
 III II
 DB 445 FKPIWR 450

RESULT 5

Q9XXP8
ID Q9XXP8 PRELIMINARY: PRT: 917 AA.
AC Q9XXP8
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE CANNA-TUBULIN RING PROTEIN DGRIP91.
CN L(1)DDA OR DGRIP91 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99156983; PubMed=10037793;
RA Oegema K., Wiese C., Martin O., Milligan R.A., Iwamatsu A.,
RA Michelson T.J., Zheng Y.;
RT "Characterization of two related Drosophila gamma-tubulin complexes
that differ in their ability to nucleate microtubules.";
RL J. Cell Biol. 144:721-733(1999).
DR EMBL: AF118380; A027817.1; .
DR Flybase: FBgn0001612; l(1)dd4.
DR InterPro: IPR000634; dehydratase_ser_thr.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 917 AA: 103819 MW: EHF2F41E35049D27 CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 6
DB 657 FKPLMR 662

RESULT 6
Q9VY94
ID Q9VY94 PRELIMINARY: PRT: 917 AA.
AC Q9VY94
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE GRIP91 PROTEIN.
CN L(1)DDA OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Colniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards M.D., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Aronson K.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Hurtis K.C., Husan D.A., Butler H., Cadieu E., Center A., Chandra L.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.H., Davies P.,
RA de Pablos H., Delcher A., Deng Z., Mays A.D., Dew C., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garq N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Hartis M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibcywan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mikulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Peinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M.D., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003493; AAF48309.1;
DR Flybase: FBgn0001612; l(1)dd4.
DR InterPro: IPR000634; dehydratase_ser_thr.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 917 AA: 103706 MW: 6AEE88C211D256BB CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 6
DB 657 FKPLMR 662

RESULT 7
Q9XJ12
ID Q9XJ12 PRELIMINARY: PRT: 221 AA.
AC Q9XJ12
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 14, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Susaki T., Nagamura Y., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
clone:PO680A03.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB023482; BAA78741.1;
SQ SEQUENCE 221 AA: 24502 MW: DC24FF370660BC15 CRC64;

Query Match 88.6%; Score 31; DB 10; Length 221;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 5
DB 602 FKPLMR 208

RESULT 3
Q9D3F2
ID Q9D3F2 PRELIMINARY: PRT: 204 AA.
AC Q9D3F2

DT 01-JUN-2001 (TREMBLREL. 17, Created)
 DT 01-JUN-2001 (TREMBLREL. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLREL. 17, Last annotation update)
 DE 5830420C20R1K PROTEIN.
 GN 5830420C20R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=THYMUS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli K., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK017942; BAH31014.1;
 DK MGD: MGI:1923275; 5830420C20R1K.
 SQ SEQUENCE 204 AA; 22892 MW; 98AF03DF7FC7B066 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 204;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 I: I I
 DB 29 FRPKWR 34

RESULT 9
 Q9YE03 PRELIMINARY; PRT; 239 AA.
 AC Q9YE03
 DT 01-NOV-1999 (TREMREL. 12, Created)
 DT 01-NOV-1999 (TREMREL. 12, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE HYPOTHETICAL 24.8 KDA PROTEIN APE0766.
 GN APE0766.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 OX NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K1;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Naqui Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 Crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).

DR EMBL: AP000060; BAA79744.1;
 DR InterPro: IPR001454; Hydrolase.
 DR InterPro: IPR000150; Hypotheset_Cof.
 DR Pfam: PF00702; Hydrolase; 1.
 DR PROSITE: PS01229; COF_2; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 239 AA; 24758 MW; EA30A3CDFD86B3DE CRC64;

Query Match 85.7%; Score 30; DB 1; Length 239;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 I: I I I
 DB 105 FRPSWQ 110

RESULT 10
 Q9D9R1 PRELIMINARY; PRT; 440 AA.
 AC Q9D9R1;
 DT 01-JUN-2001 (TREMREL. 17, Created)
 DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
 DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
 DE ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
 DE CLONE:1700030N20, FULL INSERT SEQUENCE.
 GN 5830420C20R1K.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli K., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK006565; BAB24655.1;
 DK MGD: MGI:1923275; 5830420C20R1K.
 SQ SEQUENCE 440 AA; 49566 MW; 5B5DA87E19879644 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 440;
 Best Local Similarity 66.7%; Pred. No. 3.5e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 I: I I I
 DB 39 FRPKWR 44

RESULT 11
 Q31566 PRELIMINARY; PRT; 606 AA.
 ID Q31566

AC 031566;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE YPIX PROTEIN.
 GN YPIX.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 ON NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
 RA Kobayashi Y., Koelter P., Koningstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medique C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudoga B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scollone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin H.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yanane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.P., Zumbstein K., Yoshikawa H., Danchin A.,
 KT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z99108; CAB12672.1; -;
 KW Complete proteome.
 SQ SEQUENCE 606 AA; 68677 MW; 4B58DD38922ED0CA CRC64;

Query Match 85.7%; Score 30; DB 2; Length 606;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 570 YKPEWR 575
 :|||
 :|||

RESULT 12
 ID 052961 PRELIMINARY; PRT; 610 AA.
 AC 052961;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE YPIX.
 OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 ON NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97101647; PubMed=8946165;
 RA Yamamoto H., Uchiyama S., Sekiguchi J.;
 RT "Cloning and sequencing of a 27.8-kb nucleotide sequence of the 79
 RT degrees-B1 deures region of the Bacillus subtilis genome containing
 RT the ssrE locus".
 RL DNA Res. 3:257-262(1996).
 DR EMBL: D85082; BAA24464.1; -;
 SQ SEQUENCE 610 AA; 69089 MW; 34094D05CA72FF66 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 610;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 574 YKPEWR 579
 :|||
 :|||

RESULT 13
 ID 09NA80 PRELIMINARY; PRT; 604 AA.
 AC 09NA80;
 DT 01-JUN-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
 DE Y7A10A.20 PROTEIN.
 GN Y7A10A.20.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smye R.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL117195; CAB55023.1; -;
 SQ SEQUENCE 684 AA; 79659 MW; F0B4FAF28EAF687 CRC64;

Query Match 85.7%; Score 30; DB 5; Length 684;
 Best Local Similarity 66.7%; Pred. No. 5.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 672 FKPKWK 677
 :|||
 :|||

RESULT 14
 ID 012361 PRELIMINARY; PRT; 961 AA.
 AC 012361;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE THEORETICAL 110.7 KDA PROTEIN YDL035C.
 GN GPI OR D2749 OR YDL035C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID-4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ALPHA S288C;
RA Paulin L., Saren A.M.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Paulin L., Saren A.M., Laamanen P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: 271781; CAA96454.1;
DR EMBL: 274083; CAA98593.1;
DR SGD: S0002193; GPR1.
KW Hypothetical protein.
SQ SEQUENCE 961 AA; 110708 MW; 9889D857872A4209 CRC64;

Query Match 85.7%; Score 30; DB 3; Length 961;
Best Local Similarity 66.7%; Pred. No. 7.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FKXWR 6
[1] 1:
DB 156 FKPNK 161

RESULT 15
O9LJQ1 PRELIMINARY; PRT; 1232 AA.
AC O9LJQ1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE GIAD55299.1
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC Eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID-3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX PubMed-10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
CC -1- SIMILARITY: CONTAINS A RING-TYPE ZINC FINGER.
DR EMBL: AP000414; BAB01179.1;
DR InterPro: IPR000194; APase_alpha_beta.
DR InterPro: IPR000345; CYC_heme_bind.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00097; Zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Zinc-finger.
SQ SEQUENCE 1232 AA; 139223 MW; D3770C4A9D699207 CRC64;

Query Match 85.7%; Score 30; DB 10; Length 1232;
Best Local Similarity 66.7%; Pred. No. 9.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

OY 1 FKXWR 6
[1] 1:
DB 902 FKPGWK 907

Search completed: February 27, 2002, 11:50:09
Job time: 988 sec

us-09-446-109a-7.rspt

Thu Feb 28 11:23:31 2002

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:03 ; Search time 132.19 Seconds
(without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109A-7
Perfect score: 35
Sequence: 1 FKPXR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2.6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2.6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2.6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2.6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2.6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	6	1	US-08-215-137-13
2	28	80.0	13	4	US-08-915-314-26
3	28	80.0	16	1	US-08-079-051-2
4	28	80.0	16	5	PCT-US94-06994-2
5	28	80.0	115	3	US-08-513-974B-351
6	28	80.0	243	2	US-08-829-110-3
7	28	80.0	304	1	US-08-118-270-35
8	28	80.0	304	5	PCT-US93-08528-35
9	28	80.0	350	2	US-08-458-970A-9
10	28	80.0	463	3	US-09-082-310-1
11	28	80.0	482	2	US-08-876-874-2
12	28	80.0	496	3	US-08-906-769-113
13	28	80.0	496	3	US-08-906-616-113
14	28	80.0	496	4	US-08-817-795-113
15	28	80.0	496	4	US-08-639-075A-113
16	28	80.0	496	4	US-09-012-431-113
17	28	80.0	496	4	US-09-012-692-113
18	28	80.0	496	4	US-08-906-613-113
19	28	80.0	496	5	PCT-US95-1442A-113
20	28	80.0	592	3	US-08-991-813-2
21	27	77.1	157	4	US-09-461-474-14
22	27	77.1	355	2	US-08-666-367B-6
23	27	77.1	355	4	US-09-143-438-6
24	27	77.1	448	4	US-09-461-474-8
25	27	77.1	474	4	US-09-461-474-10
26	27	77.1	527	3	US-08-907-229-2
27	27	77.1	566	2	US-08-666-367B-5

28	27	77.1	566	4	US-09-143-438-5	Sequence 5, Appli
29	27	77.1	1040	2	US-08-254-989-2	Sequence 2, Appli
30	27	77.1	1536	4	US-09-413-814-10	Sequence 10, Appl
31	27	77.1	2232	4	US-09-091-219-25	Sequence 25, Appl
32	27	77.1	2247	4	US-09-091-219-2	Sequence 2, Appli
33	26	74.3	10	2	US-08-704-655-20	Sequence 20, Appl
34	26	74.3	37	1	US-08-665-543B-4	Sequence 4, Appli
35	26	74.3	42	2	US-08-766-858A-27	Sequence 27, Appl
36	26	74.3	209	4	US-09-164-193-8	Sequence 8, Appli
37	26	74.3	273	2	US-08-997-080-75	Sequence 75, Appl
38	26	74.3	273	2	US-08-997-362-75	Sequence 75, Appl
39	26	74.3	273	3	US-08-873-970-75	Sequence 75, Appl
40	26	74.3	273	4	US-09-095-855-75	Sequence 75, Appl
41	26	74.3	336	4	US-09-334-601-13	Sequence 13, Appl
42	26	74.3	370	2	US-08-997-080-194	Sequence 194, App
43	26	74.3	370	2	US-08-997-362-194	Sequence 194, App
44	26	74.3	370	4	US-09-095-855-194	Sequence 194, App
45	26	74.3	417	1	US-08-351-981-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-215-137-13
; Sequence 13, Application US/08215137
; Patent No. 5614370
; GENERAL INFORMATION:
; APPLICANT: Konteatis, Zenon
; APPLICANT: Siciliano, Salvatore J
; APPLICANT: Springer, Martin S
; TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN C5A ANTAGONISTS
; TITLE OF INVENTION: AND AGONISTS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Ave., P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: United States of America
; ZIP: 07065

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215,137
FILING DATE:
CLASSIFICATION: N: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gerard H
REGISTRATION NUMBER: 35,746
REFERENCE/BOOKLET NUMBER: 19108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 908 594-3901
TELEFAX: 908 594-4720
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= BH
OTHER INFORMATION: /note= "either the natural phenylalanine amino
OTHER INFORMATION: terminus or the Bolton-Hunter modified peptide

OTHER INFORMATION: having the group 3-(p-hydroxyphenyl)propionyl group

FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label= dcha
OTHER INFORMATION: /note= "D-cyclohexylalanine"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /label= darg
OTHER INFORMATION: /note= "D-arginine"
US-08-215-137-13

Query Match 94.3%; Score 33; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FKPKWR 6
Db 1 FKPKWR 6

RESULT 2
US-08-915-314-26
Sequence 26, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfle, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: NO. 6180604tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

Query Match 80.0%; Score 28; DB 4; Length 13;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6

Db 4 WKPKWR 9

RESULT 3
US-08-079-051-2
Sequence 2, Application US/08079051
Patent No. 5480974
GENERAL INFORMATION:
APPLICANT: MORGAN, EDWARD L.
APPLICANT: EMBER, JULIA A.
APPLICANT: HUGLI, TONY E.
TITLE OF INVENTION: ANTIBODIES TO HUMAN C5a RECEPTOR
NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/079,051
FILING DATE: 18-JUN-1993
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Wetherell, Jr., Ph.D., John W.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: PD-2790
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: C5ar(133-148)
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
US-08-079-051-2

Query Match 80.0%; Score 28; DB 1; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 5
Db 7 FKPKWR 11

RESULT 4
US-08-06994-2
Sequence 2, Application PC/TUS9406994
GENERAL INFORMATION:
APPLICANT: The Scripps Research Institute
TITLE OF INVENTION: ANTIBODIES TO HUMAN C5a RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California

COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06994
FILING DATE: 20-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Tumarkin, Lisa A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: PD-2790
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: C5aR(133-148)
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
PCT-US94-06994-2

Query Match 80.0%; Score 28; DB 5; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 1 FKPXW 5
Db 7 FKPIW 11

RESULT 5
US-08-513-974B-351
Sequence 351, Application US/08513974B
Patent No. 6114139
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Ohtaki, Tetsuya
APPLICANT: Fukusumi, Shoji
APPLICANT: Ohgi, Kazuhiro
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
PRODUCTION, AND USE THEREOF
NUMBER OF SEQUENCES: 380
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/513.974B
FILING DATE: 14-SEP-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-093989
FILING DATE: 19-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-270017
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236357
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236356
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189274
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189273
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45753
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 351:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-513-974B-351

Query Match 80.0%; Score 28; DB 3; Length 115;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Qy 1 FKPXW 5
Db 4 FKPIW 8

RESULT 6
US-08-829-110-3
Sequence 3, Application US/08829110
Patent No. 5882890
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: NOVEL REGULATORS OF G-PROTEIN
OPERATING SYSTEM: SIGNALING
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA

COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/829,110
FILING DATE: Filed Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0259 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 243 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: THYMN0T02
CLONE: 343504
US-08-829-110-3

Query Match 80.0%; Score 28; DB 2; Length 243;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KPXR 6
DB 71 KPAWR 75

RESULT 7
US-08-118-270-35
Sequence 35, Application US/08118270
Patent No. 5508384
GENERAL INFORMATION:
APPLICANT: Murphy, Randall B.
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/118,270
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-35

Query Match 80.0%; Score 28; DB 5; Length 304;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXW 5

REFERENCE/DOCKET NUMBER: MURPHY-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-118-270-35

Query Match 80.0%; Score 28; DB 1; Length 304;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXW 5
DB 103 FKPIW 107

RESULT 8
PCT-US93-08528-35
Sequence 35, Application PC/TUS9308528
GENERAL INFORMATION:
APPLICANT: New York University
TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
NUMBER OF SEQUENCES: 348
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08528
FILING DATE: 09-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/943,236
FILING DATE: 10-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Kevin G.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 304 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-08528-35

Query Match 80.0%; Score 28; DB 5; Length 304;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXW 5

Db 103 FKPIW 107

RESULT 9

US-08-458-970A-9
: Sequence 9, Application US/08458970A
: Patent No. 5861272
: GENERAL INFORMATION:
: APPLICANT: LI, ET AL.
: TITLE OF INVENTION: C5a Receptor
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
: ADDRESSEE: CECCHI, STEWART & OLSTEIN
: STREET: 6 BECKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458, 970A
: FILING DATE: June 2, 1995
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/09234
: FILING DATE: 16 AUG 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: MULLINS, J.G.
: REGISTRATION NUMBER: 33,073
: REFERENCE/DOCKET NUMBER: 325800-353
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 350 amino acids
: TYPE: amino acid
: STRANDEDNESS: linear
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-458-970A-9

Query Match 80.0%; Score 28; DB 2; Length 350;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPW 5
Db 139 FKPIW 143

RESULT 10

US-08-082-310-1
: Sequence 1, Application US/09082310
: Patent No. 6096526
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Lal, Preeti
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl J.
: APPLICANT: Yue, Henry
: TITLE OF INVENTION: HUMAN NUCLEIC ACID METHYLASES
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 PORTER DRIVE

CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,310
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0520 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT07
CLONE: 2124957
US-09-082-310-1

Query Match 80.0%; Score 28; DB 3; Length 463;
Best Local Similarity 80.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KPXR 6
Db 293 KPSWR 297

RESULT 11
US-08-876-874-2
: Sequence 2, Application US/08876874
: Patent No. 5942405
: GENERAL INFORMATION:
: APPLICANT: Ames, Robert
: APPLICANT: Bergsma, Derk
: APPLICANT: Foley, James
: APPLICANT: Kumar, Chandrika
: APPLICANT: Naidu, Henry
: TITLE OF INVENTION: THERAPEUTIC AND SCREENING
: TITLE OF INVENTION: METHODS USING C3A RECEPTOR AND C3A
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: RATNER & PRESTIA
: STREET: P.O. BOX 980
: CITY: VALLEY FORGE
: STATE: PA
: COUNTRY: USA
: ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,874
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,627
FILING DATE: 16-JUN-1997

```

; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: P50501
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 482 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-876-874-2

Query Match      80.0%; Score 28; DB 2; Length 482;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYW 5
DB 125 FKPYW 129

RESULT 12
US-08-906-769-113
; Sequence 113, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,769
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: 24-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

; MOLECULE TYPE: protein
; US-08-876-874-2

Query Match      80.0%; Score 28; DB 3; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYW 5
DB 358 FKPYW 362

RESULT 13
US-08-906-616-113
; Sequence 113, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,616
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-906-616-113

Query Match      80.0%; Score 28; DB 3; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYW 5
DB 358 FKPYW 362

RESULT 14
US-08-817-795-113
; Sequence 113, Application US/08817795
; Patent No. 6139840
; GENERAL INFORMATION:

```

APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Frank, Glenn R.
APPLICANT: Heath, Andrew W.
APPLICANT: Yamaka, Miles Yamanaka
APPLICANT: Arfsten, Ann
APPLICANT: Dale, Beverly
APPLICANT: Stiegler, Gary
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES, AND USES THEREOF
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,795
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA: PCT/US95/14442
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gary J. Connell
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-817-795-113

Query Match 80.0%; Score 28; DB 4; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKW 5
Db 358 FKPKW 362

RESULT 15
US-08-639-075A-113
Sequence 113, Application US/08639075A
Patent No. 6150125
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-639-075A-113

Query Match 80.0%; Score 28; DB 4; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKW 5
Db 358 FKPKW 362

Search completed: February 27, 2002, 11:36:04
Job time: 143 sec

XX	21-JUL-1992.	
PD		
XX		
PF	11-DEC-1991;	91WO-US09319.
XX		
PR	27-DEC-1990;	90US-0634641.

XX This sequence represents a peptide that specifically binds to the human
 CC D2H protein. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPI1),
 CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment of
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens; (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.

XX Sequence 39 AA;

Query Match 85.7%; Score 30; DB 20; Length 39;
 Best Local Similarity 66.7%; Pred. No. 18;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKWR 6
 Db 15 frpqwr 20

RESULT 4

ID AAB01208 standard; Protein: 434 AA.

XX AAB01208;

DT 12-DEC-2000 (first entry)

DE Corn putative lecithin:cholesterol acyltransferase #3.

XX Corn; lecithin:cholesterol acyltransferase; phytosterol;
 KW phosphatidylcholine-sterol O-acyltransferase; heat shock; cold shock.

XX Zea mays.

OS WO200032791-A2.

PN 08-JUN-2000.

PF 02-DEC-1999; 99WO-US28586.

XX 03-DEC-1998; 98US-0110782.

PA (DUPO) DU PONT DE NEMOURS & CO E.L.

XX Calhoun RE, Kinney AJ, Sakai H, Shen JB, Butler RH, Saylor JJ;

XX WPI: 2000-412337/35.

DR N-PSDB: AAA49203.

XX Polynucleotide encoding plant lecithin:cholesterol acyltransferase
 PT enzyme useful for producing transgenic plants and for producing
 PT antibodies specific to which is useful for screening cDNA expression
 PT libraries

XX Claim 10; Page 40-41; 49pp; English.

XX The present sequence is a putative protein sequence of a corn
 CC lecithin:cholesterol acyltransferase (also known as
 CC phosphatidylcholine-sterol O-acyltransferase). This enzyme is found
 CC associated with high-density lipoproteins and Apolipoprotein-AI and -D.

CC The gene and protein can be used to produce transgenic plants which have
 CC increased lipid metabolism and membrane fluidity, and therefore increased
 CC resistance to heat and/or cold shock, to alter the content of phytosterol
 CC or lecithin in grains and to identify potential herbicides.

XX Sequence 434 AA;

Query Match 85.7%; Score 30; DB 21; Length 434;
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKWR 5
 Db 271 frpqwr 376

RESULT 5

ID AAG98875 standard; Protein: 462 AA.

XX AAG98875;

DT 26-SEP-2001 (first entry)

DE E. coli growth and proliferation related protein sequence SEQ ID NO:345.

XX Escherichia coli; growth; proliferation; microbial; antimicrobial;

KW Bacterial infection; microorganism.

OS Escherichia coli.

PN WO200134810-A2.

XX 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30950.

XX 09-NOV-1999; 99US-0164415.

PA (ELIPT-) ELIPTA PHARM INC.

XX Forsyth RA, Ohlsen K, Zyskind J;

XX WPI: 2001-335933/35.

DR N-PSDB: AAB84546.

XX Novel nucleic acids that inhibit Escherichia coli proliferation, useful
 PT for screening for homologous genes and for designing expression vectors

XX Claim 19; Page 421-422; 522pp; English.

XX AAU94373 to AAH84499 represent Escherichia coli growth and proliferation
 CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
 CC growth and proliferation related proteins given in AAG99078 and AAG98830
 CC to AAC98999. (I) can be used as potential targets for the generation of
 CC new antimicrobial agents, and for identification of compounds which
 CC interact with the gene products of (I). In addition the expression of
 CC (I) and the purification of the proteins, the purified proteins or other
 CC used to generate reagents and screen small molecule libraries or other
 CC candidate compound libraries for compounds that can be further developed
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
 CC complementary to (I) that are specific for particular species of
 CC microorganisms can be used to identify particular microorganism species
 CC in clinical specimens, therefore, providing a rapid and dependable
 CC method by which to identify the causative agents of a bacterial
 CC infection. Also, antibodies generated against proteins translated from
 CC mRNA transcribed from proliferation-required sequences can also be used
 CC to screen for specific microorganisms that produce such proteins in a
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing
 CC primers used in the isolation of E. coli growth and proliferation
 CC related sequence, which are used in an example from the present.

Query Match 85.7%; Score 30; DB 21; Length 847;
 Best Local Similarity 66.7%; Pred. No. 4.2e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYWR 6
 III I:
 Db 156 fkpwnk 161

RESULT 8
 AAY58358
 ID AAY58358 Standard; Protein: 864 AA.
 XX AC AAY58358;
 XX DT 27-MAR-2000 (first entry)
 XX DE Yeast G protein-coupled receptor GPR1 mutant d490-586.
 XX KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
 KW pseudohyphal form; signalling pathway; antifungal; mutant; mutin.
 XX OS Saccharomyces cerevisiae.
 OS Synthetic.
 XX PN W09963094-A2.
 XX PD 09-DEC-1999.
 XX PP 28-MAY-1999; 99WO-US11838.
 XX PR 01-JUN-1998; 98US-0088311.
 XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.
 XX PI Hirsch JP, Xue Y;
 XX P1 WPI: 2000-086980/07.
 XX PT Novel protein, gene, antibody and recombinant cell useful for
 PT identifying modulators and inhibitors useful as anti-fungals -
 XX PS Example 6.2.5; Page -: 63pp; English.
 CC The invention relates to a novel yeast G protein-coupled receptor,
 CC GPR1 (AAY58358). GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences; one
 CC sequence (KRRAQIG) is present at the N-terminal end of the loop and the
 CC other sequence (KRRAQIG) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody. GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form. The present sequence
 CC represents a GPR mutant, d277-284, used in an exemplification of
 CC the invention. This sequence is lacking the first basic motif (KRRAQIG)
 CC of the third cytoplasmic loop, relative to the native GPR1.
 CC Note: This sequence is not shown in the specification, but is derived
 CC from the yeast GPR1 amino acid sequence given in figure 1A.
 XX Sequence 864 AA;

Query Match 85.7%; Score 30; DB 21; Length 953;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;

QY 1 FKPYWR 3
 III I:
 Db 156 fkpwnk 161

RESULT 9
 AAY58356
 ID AAY58356 Standard; Protein: 953 AA.
 XX AC AAY58356;
 XX DT 27-MAR-2000 (first entry)
 XX DE Yeast G protein-coupled receptor GPR1 mutant d277-284.
 XX KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
 KW pseudohyphal form; signalling pathway; antifungal; mutant; mutin.
 XX OS Saccharomyces cerevisiae.
 OS Synthetic.
 XX PN W09963094-A2.
 XX PD 09-DEC-1999.
 XX PP 28-MAY-1999; 99WO-US11838.
 XX PR 01-JUN-1998; 98US-0088311.
 XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.
 XX PI Hirsch JP, Xue Y;
 XX P1 WPI: 2000-086980/07.
 XX PT Novel protein, gene, antibody and recombinant cell useful for
 PT identifying modulators and inhibitors useful as anti-fungals -
 XX PS Example 6.2.5; Page -: 63pp; English.
 CC The invention relates to a novel yeast G protein-coupled receptor,
 CC GPR1 (AAY58355). GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences; one
 CC sequence (KRRAQIG) is present at the N-terminal end of the loop and the
 CC other sequence (KRRAQIG) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody. GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form. The present sequence
 CC represents a GPR mutant, d277-284, used in an exemplification of
 CC the invention. This sequence is lacking the asparagine-rich region
 CC of the third cytoplasmic loop, relative to the native GPR1.
 CC Note: This sequence is not shown in the specification, but is derived
 CC from the yeast GPR1 amino acid sequence given in figure 1A.
 XX Sequence 953 AA;

Matches 4: Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPXR 6
111 1:
Db 156 fkpnrk 161

RESULT 10
AAY58357

ID AAY58357 standard: Protein; 953 AA.

XX AC AAY58357;

XX DT 27-MAR-2000 (first entry)

XX DE Yeast G protein-coupled receptor GPR1 mutant d610-617.

XX KW Yeast: G protein-coupled receptor; GPR1; filamentous growth; pseudohyphal form; signalling pathway; antifungal; mutant; mutenic.

XX OS Saccharomyces cerevisiae.

XX OS Synthetic.

XX PN W09963094-A2.

XX PD 09-DEC-1999.

XX PF 28-MAY-1999; 99WO-US11838.

XX PR 01-JUN-1998; 98US-0088311.

XX PA (MOUN) MOUNT SINAI SCHOOL MEDICINE.

XX PI Hirsch JP, Xue Y;

XX DR WPI; 2000-086980/07.

XX PT Novel protein, gene, antibody and recombinant cell useful for identifying modulators and inhibitors useful as anti-fungals -

XX PS Example 6.2.5: Page -: 63pp; English.

XX CC The invention relates to a novel yeast G protein-coupled receptor, GPR1 (AAY58357). GPR1 is predicted to have seven transmembrane domains, which is characteristic of G protein-coupled receptors, and also contains a very large third cytoplasmic loop and a large cytoplasmic tail. The third cytoplasmic loop contains two short basic sequences; one sequence (KKRAQIQ) is present at the N-terminal end of the loop and the other sequence (KKRAQIQ) is at the C-terminal end of the loop. The third cytoplasmic loop also contains an asparagine-rich region. GPR1 is believed to provide the upstream signal that activates GPA2, and plays an essential role in inducing the switch from non-filamentous to filamentous growth in yeast. Modulators of GPR1 function, especially inhibitors, are useful for prevention or treatment of a disease or disorder involving a fungal infection. The inhibitor may be an anti-GPR1 antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an inserted heterologous sequence inactivates the biological activity of GPR1. Compounds which modulate GPR1 gene expression and/or gene product activity may be used to inhibit the conversion of a fungus from a non-filamentous form to a filamentous form. The present sequence represents a GPR1 mutant, d610-617, used in an exemplification of the invention. This sequence is lacking the second basic motif (KKRAQIQ) of the third cytoplasmic loop, relative to the native GPR1. Note: This sequence is not shown in the specification, but is derived from the yeast GPR1 amino acid sequence given in figure 1A.

XX Sequence 953 AA;

Query Match 85.7%; Score 30; DB 21; Length 953;

Best Local Similarity 66.7%; Pred. NO. 4.7e+02;

Matches 4: Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 PKPXR 6
111 1:
Db 156 fkpnrk 161

RESULT 11
AAY58355
ID AAY58355 standard: Protein; 961 AA.
XX AC AAY58355;
XX DT 27-MAR-2000 (first entry)
XX DE Yeast G protein-coupled receptor GPR1.
XX KW Yeast: G protein-coupled receptor; GPR1; filamentous growth; pseudohyphal form; signalling pathway; antifungal.
XX OS Saccharomyces cerevisiae.

Key	Location/Qualifiers
FT Region	1..55
FT	/note= "Extracellular N-terminus"
FT Domain	56..80
FT	/note= "Transmembrane domain 1"
FT Region	81..90
FT	/note= "Intracellular loop 1"
FT Domain	91..115
FT	/note= "Transmembrane domain 2"
FT Region	116..132
FT	/note= "Extracellular loop 1"
FT Domain	133..156
FT	/note= "Transmembrane domain 3"
FT Region	157..180
FT	/note= "Intracellular loop 2"
FT Domain	181..200
FT	/note= "Transmembrane domain 4"
FT Region	201..246
FT	/note= "Extracellular loop 2"
FT Domain	247..276
FT	/note= "Transmembrane domain 5"
FT Misc-difference	250
FT	/note= "Encoded by AGC"
FT Region	277..620
FT	/note= "Intracellular loop 3"
FT Region	277..284
FT	/note= "Basic motif 1 (KKRAQIQ)"
FT Region	490..586
FT	/note= "Asparagine-rich region"
FT Region	610..617
FT	/note= "Basic motif 2 (KKRAQIQ)"
FT Domain	621..645
FT	/note= "Transmembrane domain 6"
FT Region	646..656
FT	/note= "Extracellular loop 3"
FT Domain	657..678
FT	/note= "Transmembrane domain 7"
FT Region	679..961
FT	/note= "Intracellular C-terminus"
FT Misc-difference	250
FT	/note= "Encoded by AGC"
XX W09963094-A2.	
XX 09-DEC-1999.	
XX 28-MAY-1999; 99WO-US11838.	
XX 01-JUN-1998; 98US-0088311.	
XX (MOUN) MOUNT SINAI SCHOOL MEDICINE.	
XX Hirsch JP, Xue Y;	

XX WPT: 2000-086980/07.
DR N-PSDR: AAZ55699.
XX Novel protein, gene, antibody and recombinant cell useful for
PT identifying modulators and inhibitors useful as anti-fungals -
XX
XX
PS Claim 4: Fig 1A: 63pp: English.
XX
XX This sequence represents a novel yeast G protein-coupled receptor,
CC GPR1. GPR1 is predicted to have seven transmembrane domains,
CC which is characteristic of G protein-coupled receptors, and also
CC contains a very large third cytoplasmic loop and a large cytoplasmic
CC tail. The third cytoplasmic loop contains two short basic sequences; one
CC sequence (KKRAQIG) is present at the N-terminal end of the loop and the
CC other sequence (KKRAQIQ) is at the C-terminal end of the loop. The
CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
CC believed to provide the upstream signal that activates GPA2, and plays
CC an essential role in inducing the switch from non-filamentous to
CC filamentous growth in yeast. Modulators of GPR1 function, especially
CC inhibitors, are useful for prevention or treatment of a disease or
CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
CC inserted heterologous sequence inactivates the biological activity of
CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
CC activity may be used to inhibit the conversion of a fungus from a
CC non-filamentous form to a filamentous form.
XX
XX Sequence 961 AA:
SQ

Query Match 85.7%; Score 30; DB 21; Length 961;
Best local Similarity 66.7%; Pred. No. 4.8e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;
QY 1 FKPKWR 6
DB 156 fkpwr 161

RESULT 12
AAR41001
ID AAR41001 standard; Protein: 1093 AA.
XX
XX AAR41001:
AC
DT 25-FEB-1994 (first entry)
XX
XX Human myotonic dystrophy gene protein.
DE
XX Abnormality: muscular dystrophy; CHR 19; chromosome 19;
KW protein kinase; polymerase chain reaction; brain.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FT Region 1..1093
FT /note= "encoded by predicted reading frame a,
FT x's in the sequence indicate stop codons
FT in the reading frame"
XX
XX W0317104-A.
XX
XX 02-SEP-1993.
XX
XX 19-FEB-1993; 93WO-US01545.
XX
XX 20-FEB-1992; 92US-0819255.
XX
XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
XX
XX Brook JD, Housman DE;
PI
XX

DR NPI: 1993-248410/36.
XX
XX DNA sequence of myotonic dystrophy gene - used to produce probes
PT and identify CHR 19 abnormality and protein kinase responsible
XX
XX Disclosure: Fig 6: 64pp; English.
XX
XX The sequence is that encoded by predicted reading frame a of
CC the human myotonic dystrophy (DM) gene. It may be used in the
CC identification of individuals affected by DM.
XX
XX Sequence 1093 AA:
SQ

Query Match 85.7%; Score 30; DB 14; Length 1093;
Best local Similarity 66.7%; Pred. No. 5.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;
QY 1 FKPKWR 6
DB 382 fkpwr 387

RESULT 13
AAY51233
ID AAY51233 standard; Protein: 2204 AA.
XX
XX AAY51233:
AC
DT 07-APR-2000 (first entry)
XX
XX Newcastle disease virus LaSota genome encoded protein 6.
DE
XX Avian-paramyxovirus; infection: lentogenic; F protein; vaccine;
KW respiratory disease; gastrointestinal disease; poultry pathogen;
KW local immunity.
XX
XX Newcastle disease virus.
OS
XX W09966045-A1.
XX
XX 23-DEC-1999.
XX
XX 17-JUN-1999; 99WO-NL00377.
XX
XX 14-JUN-1998; 98EP-0202054.
XX
XX (DIEN-) STICHTING DIENST LANDBOUWKINDIG ONDERZOE.
XX
XX Peeters BPI, De Leeuw OS, Koch G, Gielkens ALJ;
XX
XX W01: 2000-106102/09.
XX
XX New avian paramyxovirus cDNA, useful for production of vaccine against
XX Newcastle disease virus
XX
XX Disclosure: Fig 3: 115pp; English.
XX
XX This invention describes a novel avian-paramyxovirus cDNA (1) which
XX comprises a nucleic acid sequence corresponding to the 5' terminal
XX end of the genome of avian-paramyxovirus allowing the generation of
XX an infectious copy of avian-paramyxovirus. The cell line is useful for
XX the production of infectious lentogenic NIV (Newcastle disease virus)
XX without the addition of exogenous proteolytic activity. Also it is
XX possible to generate a stable transfected cell line that expresses the
XX cold-type F protein in the virus envelope therefore providing infectious
XX particles, useful in the form of a vaccine, especially against
XX respiratory and/or gastrointestinal diseases. NDV can be easily cultured
XX at very high titers in embryonated eggs. Mass culture of embryonated
XX eggs is relatively cheap. NDV vaccines are relatively stable and can be
XX simply administered by mass application methods e.g. drinking water or
XX by spraying or by aerosol formation. The natural route of infection is
XX by the respiratory and/or gastrointestinal tract which are also the major

CC routes of infection of many other poultry pathogens. NDV can induce local
CC immunity despite the presence of circulating maternal antibody. This
CC sequence represents a protein encoded by the NDV strain Lasota genome
CC which is described in the method of the invention.

XX SQ Sequence 2204 AA;

Query Match 85.7%; Score 30; DB 21; Length 2204;
Best Local Similarity 66.7%; Pred. No. 1; Indels 0; Gaps 0;
Matches 4; Conservative 1; Mismatches 1;

QY 1 FKPMR 6
I: I I
DB 1844 frptwr 1849

RESULT 14

AA053989
ID AAB53989 standard; Protein: 90 AA.

XX AC AAB53989;

XX DT 09-MAR-2001 (first entry)

XX DE Human colon cancer antigen protein sequence SEQ ID NO:1529.

XX KW Human: colon cancer: colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotropic; anti-infective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.

XX US Homo sapiens.

XX PN W0200055351-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000MO-IIS05883.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPT: 2000-587534/55.

XX DH N-PSDB; AAC98746.

XX PT Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer.

XX PS Claim 11; Page 2083-2084; 2104pp; English.

XX CC AAC9791 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnary, nephrotropic, anti-infective and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present

CC invention.

XX SQ Sequence 90 AA;

Query Match 82.9%; Score 29; DB 21; Length 90;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMR 6
I: I I
DB 74 fkcwk 79

RESULT 15

AA093531
ID AAB93531 standard; Protein: 217 AA.

XX AC AAB93531;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:12886.

XX KW Human: primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EPI074617-A2.

XX PD 07-FEB-2001.

XX PE 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Q.T. T. Isoqai T. Nishikawa T. Hayashi K. Saito K. Yamamoto J.

XX PR Ishii S. Suqiama T. Wakamatsu A. Nagai K. Otsuki T.

XX DR WPT: 2001-318749/34.

XX PS Claim 8; SEQ ID 12886; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAB03166 to AAB13628 and
CC AAB13633 to AAB18742 represent human cDNA sequences: AAB92446 to

GenCorp version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:19 ; Search time 145.23 Seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109A-8
Perfect score: 35
Sequence: 1 FKXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-08:
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	380	2 C64364	formate hydrogenly
2	33	94.3	412	2 E55146	hypothetical 47.4
3	33	94.3	498	2 H71279	probable fcmA prot
4	33	94.3	557	2 H86020	hypothetical prote
5	33	94.3	623	2 T18250	peptide transport
6	31	88.6	1035	2 S78199	probable maturase
7	30	85.7	239	2 H72667	hypothetical prote
8	30	85.7	462	2 E55131	hypothetical 47.5
9	30	85.7	462	2 H86002	probable amino aci
10	30	85.7	606	2 H69805	conserved hypothet
11	30	85.7	684	2 T31640	hypothetical prote
12	30	85.7	961	2 S67568	probable membrane
13	30	85.7	962	2 J65808	G protein-coupled
14	30	85.7	2204	1 RNZNIV	genome polyprotein
15	29	82.9	119	2 T36433	hypothetical prote
16	29	82.9	217	2 H64956	yedG protein - tsc
17	29	82.9	222	2 F5810	hypothetical prote
18	29	82.9	301	2 T24993	hypothetical prote
19	29	82.9	395	2 H69147	LPS biosynthesis R
20	29	82.9	412	2 H75484	hypothetical prote
21	29	82.9	435	2 C86340	protein F2D10.28
22	29	82.9	458	2 H82504	conserved hypothet
23	29	82.9	462	2 T15052	amino acid permeas
24	29	82.9	467	2 A59268	modulation competi
25	29	82.9	473	2 G72753	hypothetical prote
26	29	82.9	477	2 G77373	hypothetical prote
27	29	82.9	488	2 B82967	probable glucose-6
28	29	82.9	522	2 T05142	glutamate--cystein
29	29	82.9	554	1 S34607	carboxylesterase (

30	29	82.9	557	2 T39308	asparagine synthas
31	29	82.9	655	2 T30044	hypothetical prote
32	29	82.9	732	2 T50143	pumilio family pro
33	29	82.9	1224	2 S25952	gene cob intron 3
34	29	82.9	2685	2 T38755	hypothetical prote
35	28	80.0	84	2 T48944	hypothetical prote
36	28	80.0	162	2 S28301	hypothetical prote
37	28	80.0	174	2 H86400	hypothetical prote
38	28	80.0	195	2 T48945	hypothetical prote
39	28	80.0	227	2 B70438	hypothetical prote
40	28	80.0	233	1 VMWJBV	envelope protein E
41	28	80.0	254	2 S10929	transposase - Myco
42	28	80.0	254	2 A49895	transposase - Flav
43	28	80.0	273	2 T37841	probable transloca
44	28	80.0	275	2 JE0183	chitinase (EC 3.2.
45	28	80.0	297	2 E84788	hypothetical prote

ALIGNMENTS

RESULT 1
C:54364
Formate hydrogenlyase, subunit 5 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: C64364
R:Butler, G.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Ulak
son, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Accession: Kaine, 3 P.; Horodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Accession: C64364
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-380 <BUI>
A:Cross references: GB:067501; GB:L77117; NID:92826289; PIDN:AND98504.1; PID:q1591219
F:Map position: REV458767-457625
C:Superfamily: NADH dehydrogenase (ubiquinone) 49K protein

Query Match 94.3%; Score 33; DB 2; Length 380;
Best Local Similarity 83.3%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Q: 1 FKXWR 6
|||
DB 241 FKXWR 246
RESULT 2
355145
Hypothetical cal 47.4 kD protein in rsh-pit intergenic region - Escherichia coli (strain
H. Altshuler names: hypothetical protein f409
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: E55146; S47710
R:Blotauer, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Hurland, V.; Kiley, M.;
A.: 483; D.J.; 440, B.; Shao, Y.
Science 277, 1451-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: E55116
A:Status: nucleic acid sequence not shown; translation not shown.
A:Molecule type: DNA
A:Residues: 1-412 <BLAT>
A:Cross references: GB:AE000425; GB:U00096; NID:q2367232; PIDN:AC76515.1; PID:q24672
A:Experimental source: strain K-12, substrain MG1655
P:Plattner, G.
Submitted to the MHL Data Library, March 1994

A:Reference number: S47666
 A:Accession: S47710
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-398, 'MISNCACGSI' <PLUS>
 A:Cross-references: EMBL:000039; NID:g466582; PIDN:AAB18466.1; PID:g466627
 A>Note: this sequence has been corrected
 C:Genetics:
 A:Gene: yhlL

Query Match: 94.3%; Score 33; DB 2: Length 412;

Best Local Similarity 83.3%; Pred. No. 30;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III II
 DB 168 FKPDWR 173

RESULT 3

probable fcmA protein - syphilis spirochete
 C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
 C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #Lexl_change 05-Nov-1999
 C:Accession: H71279
 R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Ovi-
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD-
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
 Science 281, 375-380, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: H71279

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-498 <COL>

A:Cross-references: GB:AE001251; GB:AE000520; NID:g3323111; PIDN:AAC65773.1; PID:g3323111

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0806

Query Match 94.3%; Score 33; DB 2: Length 498;

Best Local Similarity 83.3%; Pred. No. 36;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III II
 DB 240 FKPKWR 245

RESULT 4

hypothetical protein 24888 [imported] - Escherichia coli (strain O157:H7)
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #Lexl_change 31-Mar-2001
 C:Accession: H86020
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, H.; Glasner, J.D.; Rose, D.J.; Mayhew,
 Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A83480; MUID:21074935; PMID:11206551

A:Accession: H86020

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-557 <STO>

A:Cross-references: GB:AE005174; NID:g12518181; PIDN:AAC58622.1; GSPDB:GN00145; UWGP:248

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: 24888

Query Match

94.3%; Score 33; DB 2: Length 557;

Best Local Similarity 83.3%; Pred. No. 40;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III II
 DB 189 FKPDWR 194

RESULT 5

peptide transport protein - yeast (Candida albicans)

C:Species: Candida albicans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #Lexl_change 15-Oct-1999

C:Accession: T18250

R:Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, November 1998

A:Reference number: Z18831

A:Accession: T18250

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-623 <HAR>

A:Cross-references: EMBL:AL033503; NID:el341066; PID:cl341077; PIDN:CAA22021.1

C:Genetics:

A:Gene: plr2

A:Introns: 475/3

Query Match

94.3%; Score 33; DB 2: Length 623;

Best Local Similarity 83.3%; Pred. No. 44;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III II
 DB 445 FKPKWR 450

RESULT 6

probable maturase protein, 3 - fission yeast (Schizosaccharomyces pombe) mitochondrion
 N:Alternate names: gene cob intron 1 protein
 C:Species: mitochondrion Schizosaccharomyces pombe

C:Date: 29-Jan-1998 #sequence_revision 20-Feb-1998 #Lexl_change 21-Jul-2000

C:Accession: S78199; S10070; S67378

R:Lang, B.F.

submitted to the EMBL Data Library, August 1990

A:Description: The mitochondrial genome of the fission yeast Schizosaccharomyces pombe

chizosaccharomyces pombe and Aspergillus nidulans.

A:Reference number: S78195

A:Accession: S78199

A:Molecule type: DNA

A:Residues: 1-1035 <LAN>

A:Cross-references: EMBL:X54421

A:Experimental source: strain ad7-50h

R:Lang, B.F.; Ahne, F.; Bonen, L.

J. Mol. Biol. 184, 353-366, 1985

A:Title: The mitochondrial genome of the fission yeast Schizosaccharomyces pombe. The

C:

A:Reference number: S07274; MUID:86011547

A:Accession: S10070

A:Molecule type: DNA

A:Residues: 1-1035 <LAN>

A:Cross-references: EMBL:X02819

C:Genetics:

A:Genome: mitochondrion

A:Genetic code: SGC2

C:Superfamily: cytochrome b6 homology; cytochrome b homology; plastocyanin--plastocya

C:Keywords: heme; iron; metalloprotein; mitochondrion

F:1-228/Region: cytochrome b exon 1 encoded

F:10-228/Domain: cytochrome b homology #status atypical <CHI>

F:10-210/Domain: cytochrome b6 homology <CH6>

F:222-228/Domain: plastocyanin--plastocyanin reductase 17K protein homology #status a

F:229-1035/Region: cytochrome b intron encoded

Query Match 88.6%; Score 31; DB 2: Length 1035;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III I
 DB 935 FKPSWK 940

RESULT 7
 H72667
 hypothetical:protein APE0766 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: H72667
 K:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
 A:Reference number: A72450; MUID:99310339
 A:Accession: H72667
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-239 <RAW>
 A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAA79744.1; PID:95104429
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0766
 C:Superfamily: Methanobacterium thermoautotrophicum conserved hypothetical protein MTH10

Query Match 85.7%; Score 30; DB 2: Length 239;
 Best Local Similarity 66.7%; Pred. No. 70;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III I
 DB 105 FKPSWK 110

RESULT 8
 E65131
 hypothetical 47.5 kD protein in cysG-Ltps intergenic region - Escherichia coli (strain K
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000
 C:Accession: E65131
 R:Blattner, F.R.; Mau, B.; Shao, Y.
 A.: Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: E65131
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-462 <BLAT>
 A:Cross-references: GR:000096; NID:92367215; PIDN:AAC76395.1; PID:92367216;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: ynfM
 C:Superfamily: arginine permease

Query Match 85.7%; Score 30; DB 2: Length 462;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III I
 DB 398 YKPLWR 403

RESULT 9
 B69805
 hypothetical protein ynfM - Escherichia coli (strain
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: B69805
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Polanowski, K.; Ayoda
 Nalura 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B69805
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-462 <STO>
 A:Cross-references: GR:AE005174; NID:912517998; PIDN:AAC58478.1; CSPUB:CN00145; UMG1:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ynfM

Query Match 85.7%; Score 30; DB 2: Length 462;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III I
 DB 398 YKPLWR 403

RESULT 10
 B69805
 conserved hypothetical protein ynfX - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: B69805
 R:Kunst, F.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Natur 340, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lech, J.; Hatwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 Koeller, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y.; M. Duwa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Porlete
 Klegan, M.; Kivolta, C.; Kocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
 A:Authors: Schleich, S.; Schroeter, R.; Stoffner, P.; Sekiguchi, J.; Sekowska, A.; Se
 akuch, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
 A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
 A:Reference number: A69580; MUID:98044033
 A:Accession: B69805
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-606 <KUN>
 A:Cross-references: GR:299108; GB:AL009126; NID:92633055; PIDN:CAB12672.1; PID:ej1828
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ynfX

Query Match 85.7%; Score 30; DB 2: Length 606;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 III I
 DB 378 YKPLWR 375

RESULT 11
 B61649
 hypothetical protein Y57A10A.c - Caenorhabditis elegans

C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T31640

R:Smyle, R.

submitted to the EMBL Data Library, September 1999

A:Reference number: Z21048

A:Accession: T31640

A:Status: preliminary; translated from GD/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-684 <WIL>

A:Cross-references: EMBL:AL117195; NID:e1549729; PIDN:CAH55023.1; CESP:Y57A10A.r

A:Experimental source: clone Y57A10A

C:Genetics:

A:Gene: CESP:Y57A10A.r

A:Introns: 108/3; 132/3; 149/1; 180/3; 323/2; 430/1; 530/1; 584/2; 647/1

Query Match 85.7%; Score 30; DB 2; Length 684;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYWR 6

||| |

DB 672 FKPNWK 677

RESULT 12

S67568

probable membrane protein YML035c - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein D2749

C:Species: *Saccharomyces cerevisiae*

C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 29-Oct-1999

C:Accession: S67568

R:Paulin, L.; Saren, A.M.; Lammann, P.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67560

A:Accession: S67568

A:Molecule type: DNA

A:Residues: 1-961 <PAU>

A:Cross-references: EMBL:Z74083; NID:g1431014; PID:e252988; PID:g1431015; GSPDH:GNO0004

A:Experimental source: strain S288C

C:Genetics:

A:Gene: MIPS:YDL035c

A:Map position: 4L

C:Keywords: transmembrane protein

F:57-73/Domain: transmembrane #status predicted <TM1>

F:92-108/Domain: transmembrane #status predicted <TM2>

F:140-156/Domain: transmembrane #status predicted <TM3>

F:181-197/Domain: transmembrane #status predicted <TM4>

F:256-272/Domain: transmembrane #status predicted <TM5>

F:623-639/Domain: transmembrane #status predicted <TM6>

F:662-678/Domain: transmembrane #status predicted <TM7>

Query Match 85.7%; Score 30; DB 2; Length 961;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYWR 6

||| |

DB 156 FKPNWK 161

RESULT 13

JC5808

C: protein-coupled receptor 1 - yeast (*Saccharomyces cerevisiae*)

C:Species: *Saccharomyces cerevisiae*

C:Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 19-May-2000

C:Accession: JC5808

R:Yun, C.W.; Yamaki, H.; Nakayama, K.; Yamamoto, K.; Kumagai, H.

Biochem. Biophys. Res. Commun. 240, 287-292, 1997

A:Title: G-protein coupled receptor from yeast *Saccharomyces cerevisiae*.

A:Reference number: JC5808; MUID:98049822

A:Accession: JC5808

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-962 <YUN>

C:Comment: This protein monitors the extracellular signal such as nutrition and trans

C:Genetics:

A:Gene: gpr1

F:56-74/Domain: transmembrane #status predicted <TM1>

F:91-103/Domain: transmembrane #status predicted <TM2>

F:139-157/Domain: transmembrane #status predicted <TM3>

F:180-198/Domain: transmembrane #status predicted <TM4>

F:255-273/Domain: transmembrane #status predicted <TM5>

F:622-640/Domain: transmembrane #status predicted <TM6>

F:661-679/Domain: transmembrane #status predicted <TM7>

Query Match 85.7%; Score 30; DB 2; Length 962;
Best Local Similarity 66.7%; Pred. No. 2.6e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYWR 6

||| |

DB 156 FKPNWK 161

RESULT 14

RNAZNV

genome polypeptide - Newcastle disease virus (strain Beaudette C)

N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)

C:Species: Newcastle disease virus

C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999

C:Accession: A26747

R:Yusoff, K.; Millar, N.S.; Chambers, P.; Emmerson, P.T.

Nucleic Acids Res. 15, 3961-3976, 1987

A:Title: Nucleotide sequence analysis of the L gene of Newcastle disease virus: homolog

A:Reference number: A93665; MUID:87230982

A:Accession: A26747

A:Molecule type: mRNA

A:Residues: 1-2204 <YUS>

A:Cross-references: GB:X05399; NID:g60937; PIDN:CAA28985.1; PID:g60939

C:Genetics:

A:Gene: L

C:Supernfamily: parainfluenza virus RNA-directed RNA polymerase

C:Keywords: ATP; nucleotidyltransferase

Query Match 85.7%; Score 30; DB 1; Length 2204;
Best Local Similarity 66.7%; Pred. No. 5.8e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYWR 6

||| |

DB 1844 FRPLWR 1849

RESULT 15

T36433

hypothetical protein SCF43A.10c - *Streptomyces coelicolor*

C:Species: *Streptomyces coelicolor*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T36433

R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Bartell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999

A:Reference number: Z21598

A:Accession: T36433

A:Status: preliminary; translated from GD/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-119 <SEE>

A:Cross-references: EMBL:AL096837; PIDN:CAH48897.1; GSPDH:GNO00070; SCOF43A.10c

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOF43A.10c

Query Match 82.9%; Score 29; DB 2; Length 119;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 FKPKWR 6
DB 36 FEPQWR 41

Search completed: February 27, 2002, 11:45:19
Job time: 698.9sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:45 ; Search time 78.39 seconds
(without alignments)
2.806 Million cell updates/sec

Title: US-09-446-109A-8
Perfect score: 35
Sequence: 1 FKXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	33	94.3	535	1 YHIL_ECOLI	P37629 escherichia
2	33	94.3	623	1 PTR2_CANAL	P46030 candida alb
3	31	88.6	807	1 YHIL_SCHPO	P05511 schizosacch
4	30	85.7	445	1 YHFM_ECOLI	P45539 escherichia
5	30	85.7	2204	1 KRLP_NDVR	P11205 newcastle d
6	29	82.9	217	1 YEDK_ECOLI	P76318 escherichia
7	29	82.9	452	1 SP52_MOUSE	P97364 mus musculu
8	29	82.9	477	1 Y264_SYNY3	P73436 synechocyst
9	29	82.9	514	1 GSHI_BRAJU	O23736 brassica ju
10	29	82.9	522	1 GSHI_ARATH	P46309 arabidopsis
11	29	82.9	554	1 ESTM_MOUSE	O63880 mus musculu
12	28	80.0	151	1 HSPD_BRAJA	O69241 bradyrhizob
13	28	80.0	151	1 HSPH_BRAJA	O86110 bradyrhizob
14	28	80.0	162	1 YLP6_CAEEL	Q03598 caenorhabdi
15	28	80.0	167	1 CGC6_HUMAN	Q9Y3C8 homo sapien
16	28	80.0	204	1 D5BA_LEGPN	P50024 legionella
17	28	80.0	233	1 VENV_BEV	P27904 berne virus
18	28	80.0	300	1 CYPE_DROME	Q9VJ93 drosophila
19	28	80.0	337	1 YHJD_ECOLI	P37642 escherichia
20	28	80.0	340	1 C5AR_GORGO	P79175 gorilla gor
21	28	80.0	340	1 C5AR_PANTR	P79240 pan troglod
22	28	80.0	347	1 C5AR_MOUSE	P30093 mus musculu
23	28	80.0	350	1 C5AR_HUMAN	P21730 homo sapien
24	28	80.0	352	1 C5AR_RAT	P97520 rattus norv
25	28	80.0	388	1 DXP_BACSU	O31753 bacillus su
26	28	80.0	405	1 H547_CHICK	P13731 gallus gall
27	28	80.0	417	1 H547_HUMAN	P29043 homo sapien
28	28	80.0	417	1 H547_MOUSE	P19324 mus musculu
29	28	80.0	417	1 H547_RAT	P29457 rattus norv
30	28	80.0	418	1 CRP2_HUMAN	P50454 homo sapien
31	28	80.0	469	1 RDXA_RHOSH	Q01854 rhodobacter
32	28	80.0	482	1 C3AR_HUMAN	Q16581 homo sapien
33	28	80.0	482	1 PUR8_YEAST	Q05911 saccharomyc

34	28	80.0	490	1 C883_ARATH	O21051 arabidopsis
35	28	80.0	501	1 YDF7_SCHPO	Q10479 schizosacch
36	28	80.0	508	1 GLPK_MYCLF	Q9CB81 mycobacteri
37	28	80.0	639	1 V7OK_PLRVI	P17519 potato lcaf
38	28	80.0	843	1 POL_MIYAK	P03357 akr murine
39	28	80.0	901	1 PODK_TREPA	O83728 treponema p
40	28	80.0	909	1 Y4G1_RHISN	P55465 rhizobium s
41	28	80.0	967	1 Y5G1_CLOAD	P33747 clostridium s
42	28	80.0	982	1 HBLI_CAEEL	O9XYD3 caenorhabdi
43	28	80.0	1133	1 ATX9_TETTH	O95050 tetrahymena
44	28	80.0	1196	1 POL_MLVAV	P03356 akr murine
45	28	80.0	1275	1 RFDC_MYXXA	Q50864 myxococcus

ALIGNMENT'S

RESULT 1
YHIL_ECOLI STANDARD: PRT; 535 AA.
AC P37629; P37628;
GF 01-OCT-1994 (Rel. 30, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
ET 29-AUG-2001 (Rel. 40, Last annotation update)
DE SYNTHETICAL 51.6 KDA PROTEIN IN RUSB-PIT INTERGENIC REGION.
SN YHIL OR B3H19/B3490.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision: Enterobacteriaceae;
OC Escherichia.
CX ACBT_TaxID=552;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MC1655;
RA Media: H-94316500; PubMed-8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT Analysis of the Escherichia coli genome. V. DNA sequence of the
ET Region from 75.0 to 81.5 minutes.
EL Nucleic Acids Res. 22:2576-2586(1994).
RN 121
HP REVISIONS.
RC STRAIN-K12 / MC1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Rae J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Ward B., Shao Y.;
RT The complete genome sequence of Escherichia coli K-12.
EL Nucleic Acids Res. 27:1453-1474(1997).
RN 131
HP CONCEPTUAL TRANSLATION.
RA YHIL K.F.;
RP Published observations (JUN-1999).
CC CAUTION: THIS IS A CONCEPTUAL TRANSLATION: A FRAMESHIFT WAS
INTRODUCED IN POSITION 399 TO PRODUCE THIS ORF.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC YHIL: 000019; AAB18465.1; ALT_FRAME.
LR YHIL: 000039; AAB18466.1; ALT_FRAME.
DR YHIL: AE000425; AAC76514.1; ALT_FRAME.
DR YHIL: AE000125; AAC76515.1; ALT_FRAME.
DR YHIL: E01227; YHIL.
KW Syntheical protein; Complete proteome.
SQ SEQUENCE 535 AA; 61576 MW; 0046CE539989FDF1 CRC64;

Query Match 94.3% Score 33; DB 1; Length 535;

Best Local Similarity 83.3%; Pred. No. 22;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 168 FKPMR 173

RESULT 2
 ID PTR2_CANAL STANDARD; PRT; 623 AA.
 AC P46030;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE PEPTIDE TRANSPORTER PTR2.
 GN PTR2.
 OS Candida albicans (Yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetales;
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 CC NCBI_TaxID=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 562 / ATCC 18804;
 RX MEDLINE=95291458; PubMed=7773409;
 RA Basral M.A., Lubkowitz M.A., Perry J.R., Miller D., Krainer E.,
 RA Naider F.R., Hecker J.M.;
 RT "Cloning of a Candida albicans peptide transport gene.";
 RL Microbiology 141:1147-1156(1995).
 CC -1- FUNCTION: UPTAKE OF SMALL PEPTIDE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- SIMILARITY: BELONGS TO THE PTR2 FAMILY OF TRANSPORTERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; 009781; AA80167.1; -;
 DR InterPro: IPR000109; PTR2.
 DR Pfam: PF00854; PTR2; 1.
 DR PROSITE; PS01022; PTR2.1; 1.
 DR PROSITE; PS01023; PTR2.2; 1.
 KW Peptide transport; Transport; Transmembrane.
 FT TRANSMEM 134 154
 FT TRANSMEM 163 183 POTENTIAL.
 FT TRANSMEM 191 211 POTENTIAL.
 FT TRANSMEM 250 270 POTENTIAL.
 FT TRANSMEM 277 297 POTENTIAL.
 FT TRANSMEM 385 405 POTENTIAL.
 FT TRANSMEM 418 438 POTENTIAL.
 FT TRANSMEM 448 468 POTENTIAL.
 FT TRANSMEM 499 519 POTENTIAL.
 FT TRANSMEM 529 549 POTENTIAL.
 FT TRANSMEM 557 577 POTENTIAL.
 SQ SEQUENCE: 623 AA; 69941 MW; 1601FD3AE21H80EH CRC64;

Query Match 94.3%; Score 33; DB 1; Length 623;
 Best Local Similarity 83.3%; Pred. No. 25;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 445 FKPIWR 450

RESULT 3
 ID YW01_SCIPO STANDARD; PRT; 807 AA.

P05511;
 01-NOV-1988 (Rel. 09, Created)
 15-DEC-1998 (Rel. 37, Last sequence update)
 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHEICAL 91 KDA PROTEIN IN COB INTRON.
 OS Schizosaccharomyces pombe (Fission yeast).
 CC Mitochondrion.
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC Schizosaccharomycetes.
 CC NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AD7-50;
 RX MEDLINE=86011547; PubMed=4046021;
 RA Lang B.F., Ahne F., Bonen L.;
 RT "The mitochondrial genome of the fission yeast Schizosaccharomyces
 RT pombe. The cytochrome b gene has an intron closely related to the
 RT first two introns in the Saccharomyces cerevisiae cox1 gene.";
 RL J. Mol. Biol. 184:353-366(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AD7-50;
 RA Lang B.F.;
 RL Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS CODED IN THE GROUP-II INTRON OF
 CC COB.
 CC -1- SIMILARITY: TO GROUP II INTRON MATURASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X02819; CAA26587.1; -;
 DR EMBL; X54421; CAA38288.1; -;
 DR InterPro: IPR003615; HNH_nuc.
 DR InterPro: IPR000442; Intron_maturase2.
 DR InterPro: IPR000477; RVTse.
 DR Pfam: PF01348; Intron_maturase2; 1.
 DR Pfam: PF00078; rvt; 1.
 DR SMART; SM00507; HNHc; 1.
 KW Hypothetical protein; Mitochondrion.
 SQ SEQUENCE 807 AA; 91637 MW; AFD70F35E84CAABF CRC64;

Query Match 88.6%; Score 31; DB 1; Length 807;
 Best Local Similarity 66.7%; Pred. No. 78;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 DB 707 FKPSWK 712

RESULT 4
 ID YHFM_ECOLI STANDARD; PRT; 445 AA.
 AC P45339; P76686;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHEICAL 47.5 KDA PROTEIN IN CYSG-TRPS INTERGENIC REGION.
 GN YHFM OR B3370.
 OS Escherichia coli.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 CC NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT *The complete genome sequence of Escherichia coli K-12.*;
 RL Science 277:1453-1474(1997).
 CC -1- SURCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE AKIC/CADH/GERAB/LYSI FAMILY OF
 CC PERMEASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U18997; AAC58167.1; ALT_SEQ.
 DR EMBL: AK000413; AAC76395.1; ALT_INIT.
 DR EcoGene: EGI2908; yhm.
 DR InterPro: IPR002293; AA_rel_permease-1.
 DR InterPro: IPR002027; Amino_acid_permease.
 DR Pfam: PF00324; aa_permeases; 1.
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 38 58 POTENTIAL.
 FT TRANSMEM 93 113 POTENTIAL.
 FT TRANSMEM 121 141 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 181 201 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 FT TRANSMEM 273 293 POTENTIAL.
 FT TRANSMEM 334 354 POTENTIAL.
 FT TRANSMEM 355 375 POTENTIAL.
 FT TRANSMEM 389 410 POTENTIAL.
 FT TRANSMEM 417 435 POTENTIAL.
 SQ SEQUENCE 445 AA: 47576 MW: 89805944.45E1PEF CRC64;

 Query Match 85.7%; Score 30; DB 1; Length 445;
 Best Local Similarity 66.7%; Pred. No. 71;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FRPXWR 6
 DB 381 YKPLWR 386
 RESULT 5
 RRPL_NDVB STANDARD: PRT: 2204 AA.
 AC P11205;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
 DE (L PROTEIN).
 GN L.
 OS Newcastle disease virus (strain Heaudette C/45) (NIIV).
 OC Viruses: ssRNA negative-strand viruses; Mononegavirales;
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirus.
 OX NCBI_TaxID=11178;
 RN [1]
 RI MEDLINE=87230482; PubMed=3035486;
 RX Yusoff K., Miller N.S., Chambers P., Kimmerman P.T.;
 RA *Nucleotide sequence analysis of the L gene of Newcastle disease
 RT virus: homologues with Sendai and vesicular stomatitis viruses.*;

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

RL Nucleic Acids Res. 15:3961-3976(1987).
 CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
 CC SYNTHESIZE IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)
 CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs, RNA EDITING OF THE P
 CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
 CC -1- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X05339; CAA28985.1;
 DR PIR: A26747; RRN2NV.
 DR InterPro: IPR001016; Paramyx_RNA_pol.
 DR Pfam: PF00946; Paramyx_RNA_pol; 1.
 KW Transferase; RNA-directed RNA polymerase.
 SQ SEQUENCE 2204 AA: 248622 MW: C67B8674D904802C CRC64;

 Query Match 85.7%; Score 30; DB 1; Length 2204;
 Best Local Similarity 66.7%; Pred. No. 3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FRPXWR 6
 DB 1844 FRPLWR 1849
 RESULT 5
 YEDK_ECOLI STANDARD: PRT: 217 AA.
 ID YEDK_ECOLI
 AC P75418;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 21-AUG-2001 (Rel. 40, Last annotation update)
 DE SYNTHETICAL, 24.5 KDA PROTEIN IN AMYA-F11E INTERGENIC REGION.
 GN YEDK OR B1931.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterolacteriaceae;
 OC Escherichia.
 OX ACBI_TaxID=562;
 RN [1]
 RI SEQUENCE FROM N.A.
 RC STRAIN=JALL;
 XX MEDLINE=933d1452; PubMed=8371104;
 RA Baba M., Kihara M., Kawagishi I., Marnab R.M.;
 RT Topanization of the Escherichia coli and Salmonella typhimurium
 RT chromosomes between flagellar regions IIIa and IIIb, including a
 RT large non-coding region.*;
 RL J. Gen. Microbiol. 139:1401-1407(1993).
 RN [2]
 RI SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT *The complete genome sequence of Escherichia coli K-12.*;
 RL Science 277:1453-1474(1997).
 CC -1- SIMILARITY: TO YEAST YMR114C.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC EMBL: L13279; ; NOT_ANNOTATED_CDS.
DR EMBL: AF000285; AAC74998.1; -.
DR EcoGeno: BG13278; ycdK.
DR InterPro: IPR003738; DUF159.
DR Pfam: PF02586; DUF159; 1.
KW HYPOTHETICAL PROTEIN: Complete proteome.
FT CONFLICT 213 217 TRASN -> NOGAELIOPV (IN REF. 1).
SQ SEQUENCE 217 AA; 24500 MW; 4F26C95DB3H02900 CRC64;

Query Match: 82.9%; Score 29; DB 1; Length 217;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 4; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 FKPKWR 6
Db 87 FKPLWQ 92

RESULT 7
SPS2_MOUSE STANDARD; PRT; 452 AA.
AC P97364;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SELENIDE WATER KINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)
GN SPS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN 1;
RS SEQUENCE FROM N.A.
RA MEDLINE=96017645; PubMed=7588067;
RA Guimaraes M.J., Bazan J.F., Zlotnik A., Willes M.V., Grimaldi J.C.,
RA Lee F., McClanahan T.;
RA "A new approach to the study of haematopoietic development in the
RA yolk sac and embryoid bodies";
RL Development 121:3335-3346(1995).
RW [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97140286; PubMed=8986768;
RA Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Terriek D.A., Kastelein R., Bazan J.F.,
RA Zlotnik A.;
RT "Identification of a novel self homolog from eukaryotes, bacteria,
RT and archaea: is there an autoregulatory mechanism in selenocysteine
RT metabolism?";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).
CC -!- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.
CC -!- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O -> AMP + SELENOPHOSPHATE
CC + PHOSPHATE.
CC -!- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
CC ENCODED BY THE OPAL CODON, UGA.
CC -!- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.
CC CLASS 1 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U43285; AAC53024.1; -.
CC MOP; MCI:108388; SPS2.
DR InterPro: IPR000728; AIRS_related.
DR Pfam: PF00586; AIRS; 1.
KW Transferrase; Selenium; Selenocysteine; ATP-binding.

```

```

FT ACT_SITE 63 63 POTENTIAL.
FT SE_CYS 63 63
FT SITE 66 66 IMPORTANT FOR CATALYTIC ACTIVITY (HY
FT SIMILARITY).
FT NP_BIND 322 328 ATP (POTENTIAL).
FT DOMAIN 2 9 POLY-ALA.
FT DOMAIN 433 440 POLY-ALA.
SQ SEQUENCE 452 AA; 47786 MW; 9DA6F7250CFE80E4 CRC64;

Query Match: 82.9%; Score 29; DB 1; Length 452;
Best Local Similarity 66.7%; Pred. No. 11e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FKPKWR 6
Db 48 FSPSWR 53

RESULT 8
Y264_SYNY3
ID Y264_SYNY3 STANDARD; PRT; 477 AA.
AC P73436;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 54.0 KDA PROTEIN SLL1464.
GN SLL1464.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN 1;
RS SEQUENCE FROM N.A.
RA MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RA "Sequence analysis of the genome of the unicellular cyanobacterium
RA Synecocystis sp. strain PCC6803. II. Sequence determination of the
RA entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D90906; BAA17476.1; -.
DR InterPro: IPR003846; UPF0061.
DR Pfam: PF02696; UPF0061; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 477 AA; 54041 MW; 81F6899B1A6D613C CRC64;

Query Match: 82.9%; Score 29; DB 1; Length 477;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 FKPKWR 6
Db 398 FSPSWR 393

RESULT 9
CSHI_BRAJU
ID GSH1_BRAJU STANDARD; PRT; 514 AA.
AC O21736; Q43389;

```

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE: GLUTAMATE--CYSTEINE LIGASE, CHLOROPLAST PRECURSOR (EC 6.3.2.2) (GAMMA-GLUTAMATE--CYSTEINE SYNTHETASE) (GAMMA-ECS) (GCS).
 DE: GLUTAMATE--CYSTEINE LIGASE, CHLOROPLAST PRECURSOR (EC 6.3.2.2) (GAMMA-GLUTAMATE--CYSTEINE SYNTHETASE) (GAMMA-ECS) (GCS).
 GN: Brassica juncea (Leaf mustard) (Indian mustard).
 OS: Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC: Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC: eucosids II; Brassicales; Brassicaceae; Brassica.
 OX: NCBI_TaxID=3707;
 RN: [1]
 RP: SEQUENCE FROM N.A.
 RC: TISSUE=RCOL;
 RX: MEDLINE=98281577; PubMed=9620267;
 KX: Schaefer H.J., Haag-Korner A., Rausch T.;
 KT "cDNA cloning and expression analysis of genes encoding GSH synthesis in roots of the heavy-metal accumulator Brassica juncea L.: evidence for Cd-induction of a putative mitochondrial gamma-glutamylcysteine synthetase isoform";
 RT Plant Mol. Biol. 37:87-97(1998).
 RL [2]
 RN SEQUENCE OF 126-342 FROM N.A.
 RP STRAIN=CV. VITTASSO; TISSUE=Leaf;
 RC MEDLINE=97227950; PubMed=9119067;
 RX Schaefer H.J., Greiner S., Rausch T., Haag-Korner A.;
 RT "In seedlings of the heavy metal accumulator Brassica juncea, Cu2+ differentially affects transcript amounts for gamma-glutamylcysteine synthetase (gamma-ECS) and metallothionein (MT2).";
 RT FEBS Lett. 404:216-220(1997).
 RL CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + L-CYSTEINE = ADP + ORTHOPHOSPHATE + GAMMA-L-GLUTAMYL-L-CYSTEINE.
 CC -1- PATHWAY: FIRST AND RATE-LIMITING STEP IN DE NOVO GLUTATHIONE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).
 CC -1- MISCELLANEOUS: IN PLANTS, THERE ARE TWO FORMS OF THIS ENZYME, ONE CYTOSOLIC AND THE OTHER CHLOROPLASTIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y10848; CAA71801.1;
 DR EMBL: X95563; CAA64808.1;
 KW Glutathione biosynthesis; Ligase; Chloroplast; Transit peptide.
 FT TRANSIT 1 55 CHLOROPLAST (POTENTIAL).
 FT CHAIN 56 514 GLUTAMATE--CYSTEINE LIGASE.
 FT CONFLICT 139 139 K -> R (IN REF. 2).
 FT CONFLICT 215 215 T -> I (IN REF. 2).
 FT CONFLICT 257 257 M -> T (IN REF. 2).
 SQ SEQUENCE 514 AA: 57903 MW: 07C71CB13E785FA8 CRC64;

 Query Match 82.9% Score 29 DB 1: Length 514;
 Best Local Similarity 66.7% Pred. No. 1.3e+02;
 Matches 4: Conservative 1: Mismatches 1: Indels 0: Gaps 0;

 QY 1 FKPKWR 6
 I:I:I:I
 DB 204 FKPKWR 209

 RESULT 10
 GSHL_ARATH
 ID: GSHL_ARATH STANDARD: PKT: 522 AA.
 AC: P46309; 082759; 192951;
 DT: 01-NOV-1995 (Rel. 32, Created)
 DT: 20-AUG-2001 (Rel. 40, Last sequence update)
 DT: 20-AUG-2001 (Rel. 40, Last annotation update)

DE DE
 DE DE
 DE GS1 OR AT4G23100 OR F7H19.290.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC: Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC: eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX: NCBI_TaxID=102;
 RN: [1]
 RP: SEQUENCE FROM N.A.
 RC: STRAIN=CV. COLUMBIA; TISSUE=Leaf;
 RX: MEDLINE=95023984; PubMed=7937837;
 KX: May M.J., Leaver C.J.;
 KT "Arabidopsis thaliana gamma-glutamylcysteine synthetase is structurally unrelated to mammalian, yeast, and Escherichia coli homologs";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:10059-10063(1994).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. LANDSBERG ERRECTA;
 RC Wilmann P., Gondet L., Bach T.J.;
 RT "Isolation of an Arabidopsis thaliana cDNA encoding a putative gamma-glutamylcysteine synthetase by complementation of a GSH1 deficient yeast mutant-glutamylcysteine synthetase";
 RT Submitted (JSC-1996) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. COLUMBIA;
 KX Cobbett C.S., May M.J., Howden R., Rolis B.;
 KT "The glutathione-deficient, cadmium-sensitive mutant, cad3-1, of Arabidopsis thaliana is deficient in gamma-glutamylcysteine synthetase";
 RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RL [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CV. COLUMBIA; PubMed=10617198;
 KX Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G., Wijn T., Buesterhoft A., Stekema W., Ktiani K.-D., Torry N., Harris B., Anstorge W., Brandt P., Grivell L.A., Kieger M., Weishegartner M., de Simone V., Oberwiler H., Machle K., Mueller M., Kreis M., Deiseny M., Puigdomenech P., Watson M., Schmidheini T., Reichert B., Portelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Hohseil J., Zimmermann W., Wedler H., Ridley P., Langham S.A., McCullagh B., Bitham L., Robben J., Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F., Aerts K., Weijens L., Voet M., Bastiaens I., Aert R., Defoor K., Melzer E., Brandt A., Pelers S., van Staveren M., Dickse W., Moerman P., Klein Lankhorst R., Rose M., Hauf J., Koeller P., Bernisier S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H., De Keyser A., Huysschaert C., Gleien J., Villarroel K., De Clercq R., Van Montagu M., Rogers J., Cronin A., Quail M., Hray-Allen S., Clark L., Dongett J., Hall S., Kay M., Lennard N., McIwain K., Mayes R., Pothol A., Rajandream M.-A., Lyne M., Denes V., Reumann S., Dorkova D., Abocker H., Scharle M., Grimm M., Loehner T.-H., Kase S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S., Jabit C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A., Reumann S., Argirou A., Vitale D., Liguori R., Piravandi E., Bassinet O., Quigley F., Clabaud G., Muendlein A., Felber R., Schabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S., Chastor F., Cooke R., Berger C., Montfort A., Casacuberta E., Ribbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A., Ribbons T., Purnelle H., Bent E., Johnson S., Tacon D., Jesse T., Heinen L., Schwarz S., Scholler P., Heber S., Francis P., Hielke C., Fishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S., Zaccaria P., Devan M., Wilson R.K., de la Bastide M., Habermann K., Purnell L., Dethia N., Gnoj L., Schutz K., Huang E., Spiedel L., Sakon M., Murray J., Sheet P., Cordes M., Abu-Threideh J., Sackling P., Kalicki J., Graves T., Harmon G., Edwards J., Sackling P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Sak P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Telfer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,

RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antonoli H., Zidanic M., Strong C., Sun H., Lemar H., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Iasegawa A., Iameed A., Lodhi M., Johnson A.,
 RA Chen E., Marce M., Martienssen R., McCombie W.R.,
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RT thaliana".
 RL Nature 402:769-777(1999).
 CC -- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + L-CYSTEINE - ADP +
 CC ORTHOPHOSPHATE + GAMMA-L-GLUTAMYL-L-CYSTEINE.
 CC -- ENZYME REGULATION: FEEDBACK INHIBITION BY GLUTATHIONE.
 CC -- PATHWAY: FIRST AND RATE-LIMITING STEP IN DE NOVO GLUTATHIONE
 CC BIOSYNTHESIS.
 CC -- SURUNIT: MONOMER (PROBABILE).
 CC -- SUBCELLULAR LOCATION: CHLOROPLAST (PROBABILE).
 CC -- TISSUE SPECIFICITY: ABUNDANT IN LEAVES AND ROOTS.
 CC -- ENZYME REGULATION: IN PLANTS, THERE ARE TWO FORMS OF THIS ENZYME, ONE
 CC CYTOSOLIC AND THE OTHER CHLOROPLASTIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z29490; CAA82626.1; -
 DR EMBL: Y09944; CAA71075.1; -
 DR EMBL: AF068269; AAD14544.1; -
 DR EMBL: AF031018; CAA19826.1; -
 DR EMBL: AF161558; CAA79265.1; -
 DR Mendel: 17259; Arath:2621;17259.
 DR Mendel: 33329; Arath:2621;33329.
 DR Mendel: 36232; Arath:2621;36232.
 KW Glutathione biosynthesis; Ligase; Chloroplast; Transit peptide.
 FT TRANSIT 1 ? CHLOROPLAST (POTENTIAL).
 FT CHAIN ? 522 GLUTAMATE--CYSTEINE LIGASE.
 FT VARIANT 22 22 A -> T (IN STRAIN LANDSBERG ERCTA).
 FT VARIANT 48 48 Y -> N (IN STRAIN LANDSBERG ERCTA).
 FT CONFLICT 491 522 YRTGVTPAEKLEMYNGEWSQVDFVEELLY -> GGNRS
 FT YACGEALGDVQWRMGTRRSRVRAAVALRKMDVKNRCL
 FT (IN REF. 1).
 SQ SEQUENCE 522 AA; 58562 MW; CCHF13C6F44E0E7 CRC64;
 Query Match 82.9%; Score 29; DB 1; Length 522;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKXWR 6
 DB 212 FQPKWR 217
 RESULT 11
 ESTM_MOUSE STANDARD; PRT: 554 AA.
 AC Q63880;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE LIVER CARBOXYL ESTERASE PRECURSOR (EC 3.1.1.1) (ES-MALE) (ESTERASE-31).
 GN ES31.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCHI_TaxID=10090;
 RN 111
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-DALD/CJ X DDA/2J; TISSUE-Liver;

KX MEDLINE-93326638: PubMed-7916639;
 RA Alia K., Moore R., Negishi M.;
 RT "Cloning and nucleotide sequence of a novel, male-predominant
 RT carboxylesterase in mouse liver".
 RL Biochim. Biophys. Acta 1174:72-74(1993).
 CC -- FUNCTION: INVOLVED IN THE DETOXIFICATION OF XENOBIOTICS AND IN
 CC THE ACTIVATION OF ESTER AND AMIDE PRODRUGS.
 CC -- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O - AN ALCOHOL
 CC + A CARBOXYLIC ANION.
 CC -- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE, LUMEN OF ENDOPLASMIC
 CC RETICULUM.
 CC -- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S64130; AAR27606.1; -
 DR HSP: P21836; IMAH.
 DR MGD: MGI:102773; ES31.
 DR InterPro: IPR002018; Carboxylesterase-B.
 DR InterPro: IPR000379; Est_lip_thioest_actsite.
 DR Pfam: PF00135; Coesterase; 1
 DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 DR PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
 KW Glycoprotein; Hydrolase; Serine esterase; Endoplasmic reticulum;
 KW Signal: Multigene family.
 FT SIGNAL 1 14 BY SIMILARITY.
 FT CHAIN 15 554 LIVER CARBOXYLESTERASE.
 FT ACT_SITE 215 215 BY SIMILARITY.
 FT ACT_SITE 443 443 BY SIMILARITY.
 FT DISULFID 83 110 BY SIMILARITY.
 FT DISULFID 267 278 BY SIMILARITY.
 FT SITE 551 554 PREVENT SECRETION FROM ER (POTENTIAL).
 FT CARBOHYD 294 294 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 300 300 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 381 381 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 554 AA; 61509 MW; 028D98B06F3CAA0 CRC64;
 Query Match 82.9%; Score 29; DB 1; Length 554;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKXW 5
 DB 434 FKPAW 438
 RESULT 12
 HSPD_BRAJA STANDARD; PRT: 151 AA.
 AC Q69241;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPD.
 GN HSPD.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Bradyrhizobium group; Bradyrhizobium.
 CX NCBI_TaxID=375;
 RN 111
 RP SEQUENCE FROM N.A.
 RA Natterhaus F., Weighlofer W., Fischer H.M., Hennecke H.;
 RT "Identification of the Bradyrhizobium japonicum dcp gene as part of
 RT an operon containing small heat shock protein genes".
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 CC -- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)

```

CC -----
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC FMBL; AJ003064; CAA05835.1; -.
CC InterPro: IPR002068; Crystallin_HSP20.
CC Pfam: PF00011; HSP20; 1.
CC PROSITE: PS01031; HSP20; 1.
CC Heat shock; Multigene family.
CC SEQUENCE 151 AA: 17272 MW: 17448994.959994 CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 151;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKPKXR 6
DB 6 FSLPLR 11

RESULT 13
HSPH_BRAJA STANDARD; PRT: 151 AA.
AC OR6110;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SMALL HEAT SHOCK PROTEIN HSPH.
GN HSPH.
OS Bradyrhizobium japonicum.
CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN 11
RP SEQUENCE FROM N.A.
RA Muenchbach M., Nocker A., Narberhaus F.;
RA "Occurrence of a superfamily of small heat shock proteins in
RA Bradyrhizobium japonicum and other Rhizobium species: a plant-like
RA phenomenon.";
RA Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
KW SMALL HEAT SHOCK PROTEIN (HSP20)
CC
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC FMBL; AJ010144; CAA09014.1; -.
CC InterPro: IPR002068; Crystallin_HSP20.
CC Pfam: PF00011; HSP20; 1.
CC PROSITE: PS01031; HSP20; 1.
CC Heat shock; Multigene family.
CC SEQUENCE 151 AA: 17098 MW: 17498093.1047628 CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 151;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKPKXR 6
DB 6 FSLPLR 11

```

```

RESULT 14
YLF6_CAEEL STANDARD; PRT: 162 AA.
AC O04538;
DT 01-FEB-1994 (Rel. 28, Created)
DT 31-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 18.5 KDA PROTEIN C40H1.6 IN CHROMOSOME III.
GN C40H1.6.
OS Caenorhabditis elegans.
CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabdilloidea;
CC Rhabdillidae; Peloderiinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M., Cooper A.,
RA Hontfield J., Hurton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laisster N.,
RA Laureille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaildon N., Smith A., Smith M., Sonnenhammer E., Staden K.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RC Nature 368:32-38(1994).
CC -!- SIMILARITY: STRONG, TO HUMAN CGI-126.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC FMBL; Z19154; CAA79557.1; -.
CC PIR: S28301; S28301.
CC WormPep: C40H1.6; CE00114.
KW Hypothetical protein.
SQ SEQUENCE 162 AA: 18537 MW: 80030363BCE79D55 CRC64;

Query Match
Best Local Similarity 80.0%; Score 28; DB 1; Length 162;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKXR 5
DB 120 FKPLW 124

RESULT 15
CSC6_HUMAN STANDARD; PRT: 167 AA.
AC Q9V3C8;
DT 19-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN CGI-126.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.

```

RX MEDLINE=20272150; Pubmed=10810093;
 RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
 RT "Identification of novel human genes evolutionarily conserved in
 RL Caenorhabditis elegans by comparative proteomics";
 CC Genome Res. 10:703-713(2000).
 CC -!- SIMILARITY: STRONG, TO C.ELEGANS C40H1.6.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF151084; AAD34121.1; -
 KW hypothetical protein.
 SQ SEQUENCE 167 AA: 19458 MW: 167509187DC43E14 CRC64;

Query Match 80.0% Score 28; DB 1; Length 167;
 Best Local Similarity 80.0% Pred. No. 70;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 FKXW 5
 Db 121 FKPLW 125

Search completed: February 27, 2002, 11:42:45
 Job time: 544 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:09 : Search time 281.76 Seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109A-8
Perfect score: 35
Sequence: 1 FKXWR 6

Scoring table: HUSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 14627329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREML_17:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_orcanelle:
9: sp_phage:
10: sp_plant:
11: sp_rodent:
12: sp_virus:
13: sp_vertebrate:
14: sp_unclassified:

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	34	97.1	250	2 Q9EUM2	Q9EUM2 corynebacte
2	33	94.3	380	1 Q57935	Q57935 methanococ
3	33	94.3	498	2 Q83784	Q83784 treponema p
4	33	94.3	623	3 Q9UR17	Q9UR17 candida alb
5	33	94.3	917	5 Q9XYP8	Q9XYP8 drosophila
6	33	94.3	917	5 Q9VY94	Q9VY94 drosophila
7	31	88.6	221	10 Q9XJ12	Q9XJ12 oryza sativ
8	30	85.7	204	11 Q9D3F2	Q9D3F2 mus musculu
9	30	85.7	239	1 Q9YK03	Q9YK03 aeropyrum p
10	30	85.7	440	11 Q9D9R1	Q9D9R1 mus musculu
11	30	85.7	606	2 Q31566	Q31566 bacillus su
12	30	85.7	610	2 Q52961	Q52961 bacillus su
13	30	85.7	684	5 Q9NA80	Q9NA80 caenorhabdi
14	30	85.7	961	3 Q12361	Q12361 saccharomyc
15	30	85.7	1232	10 Q91JQ1	Q91JQ1 arabidopsis
16	30	85.7	2204	12 Q90341	Q90341 newcastle d
17	30	85.7	2204	12 Q9MMH6	Q9MMH6 newcastle d
18	30	85.7	2204	12 Q9DL03	Q9DL03 newcastle d
19	30	85.7	2454	3 Q9UVP2	Q9UVP2 emerlicella

Q9UVP2 emerlicella
Q9XAA5 streptomyce
Q9ZEW4 klebsiella
Q9G402 plasmodium
Q9CJP0 pasteurella
Q9QJY1 giardia lam
Q9N2Y6 caenorhabdi
Q912R9 streptomyce
Q9FS78 triticum ae
Q9NND2 icishmania
Q33353 mycobacteri
Q22568 caenorhabdi
Q26470 methanobact
Q9A710 caulobacter
Q9RW53 deinococcus
Q91M75 arabidopsis
Q82665 brassica ju
Q9C1B2 gibberella
Q94088 fusarium sp
Q9C1B9 fusarium sp
Q9KN89 vibrio chol
Q40414 nicotiana s
Q52729 rhizobium e
Q9YGB1 aeropyrum p
Q9HTC7 pseudomonas
Q42902 schizosacch

ALIGNMENTS

RESULTS: 1
Q9EUM2
ID Q9EUM2 PRELIMINARY: PRT: 250 AA.
AC Q9EUM2
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE YCGLC.
GN YCGLC.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC plasmid R-plasmid pCG4.
OC Bac-eria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX ACRI_TaxID=1718;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC31830;
RA Tsuch A., Puhler A., Kalinowski J.;
RT "DNA sequence and genetic organization of the integron-carrying R-
plasmid pCG4 of Corynebacterium glutamicum.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF164956; AAG00294.1;
DR FaterPro: IPR002145; CopG_HTH_4.
DR EMBL: P01402; HTH_4; 1.
KW plasmid.
SQ SEQUENCE 250 AA; 28804 MW; 195C86367B483157 CRC64;

Query Match 97.1%; Score 34; DB 2; Length 250;
Bes. Local Similarity 83.3%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 FKXWR 5
OB 214 FKXWR 219

RESULTS: 2
Q57935 PRELIMINARY: PRT: 380 AA.
V: Q57935;

Q9VY98 Q9VY98 PRELIMINARY: PRT: 917 AA.

AC Q9VY98: 01-NOV-1999 (TREMELREL. 12, Created)

DT 01-NOV-1999 (TREMELREL. 12, Last sequence update)

DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)

DE GAMMA-TUBULIN RING PROTEIN DGRIP91.

GN L(1)DD4 OR GGRIP91 OR CG10988.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE=99156983; PubMed=10037793;

RA Cegennia K., Wiesse C., Martin O., Milligan R.A., Iwamatsu A.,

RA Mitchison T.J., Zheng Y.

RT "Characterization of two related Drosophila gamma-tubulin complexes

RT that differ in their ability to nucleate microtubules.";

KL J. Cell Biol. 144:721-733(1999).

DR EMBL: AF118380; AAD27817.1; ..

DR FlyBase: FBgn0001612; l(1)dd4.

DR InterPro: IPR000634; dehydratase_ser_thr.

DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.

SQ SEQUENCE 917 AA; 103819 MW; EBF2F41E35049D27 CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;

Best Local Similarity 83.3%; Pred. No. 1.9e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 6

DB 657 FKPLWR 662

RESULT 6

Q9VY94 Q9VY94 PRELIMINARY: PRT: 917 AA.

AC Q9VY94: 01-MAY-2000 (TREMELREL. 13, Created)

DT 01-MAY-2000 (TREMELREL. 13, Last sequence update)

DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)

DE GRIP91 PROTEIN

GN L(1)DD4 OR CG10988.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Anantides P.C., Scher S.K., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.C., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abrell J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Hallig R.M., Basu A., Haxendale J., Hayraktaroglu L., Heasley E.M.,

RA Beeson K.Y., Benos P.V., Bertram R.P., Bhandari D., Bolshakov S.,

RA Borkova D., Hotchan M.K., Houck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Hodson K., Hou P., Howes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Burbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gilbert W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Joslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA McKulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

OR EMBL: AE003493; AAF48309.1; ..

DR FlyBase: FBgn0001612; l(1)dd4.

DR InterPro: IPR000634; dehydratase_ser_thr.

DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.

SQ SEQUENCE 917 AA; 103706 MW; 6AE886C211D256BB CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;

Best Local Similarity 83.3%; Pred. No. 1.9e+02;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 3

DB 657 FKPLWR 662

RESULT 7

Q9XJ12 Q9XJ12 PRELIMINARY: PRT: 221 AA.

AC Q9XJ12: 01-NOV-1999 (TREMELREL. 12, Created)

DT 01-NOV-1999 (TREMELREL. 12, Last sequence update)

DT 01-JUN-2000 (TREMELREL. 14, Last annotation update)

DE HYPOTHETICAL PROTEIN.

OS Oryza sativa (Rice).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Zingiberaceae; Oryzae; Oryza.

OX NCBI_TaxID=4530;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. NIPPONBARE;

RA Raseki T., Nagamura Y., Yamamoto K.;

RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC

RT clone: P0680A03.";

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB023482; BAA7841.1; ..

SQ SEQUENCE 221 AA; 24502 MW; DC24FF370660BC15 CRC64;

Query Match 88.6%; Score 31; DB 10; Length 221;

Best Local Similarity 66.7%; Pred. No. 1.2e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 6

DB 203 FKPLWR 208

RESULT 8

Q9D3F2 Q9D3F2 PRELIMINARY: PRT: 204 AA.

AC Q9D3F2

DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE 5830420C20R1K PROTEIN.
 CN 5830420C20R1K.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=THYMUS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Hatalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Hoffelli D., Bojunga N., Carninci P., Fujita M., Gariboldi M.,
 RA Brownstein M.J., Hult C., Fletcher C., Carninci P., de Bonaldo M.F.,
 RA Guslinich S., Hill D., Holmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK017942; HAB31014.1;
 DR MGI: 1923275; 5830420C20R1K.
 SO SEQUENCE 204 AA; 2292 MW; 99A03DF7C7FDB066 CRC64;

Query Match 85.7%; Score 30; DH 11; Length 204;
 Hest Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKXWR 6
 DB 29 FRPRMR 34
 1:1 11

RESULT 9
 QYVE03 PRELIMINARY; PRT; 239 AA.
 AC QYVE03;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE HYPOTHETICAL 24.8 KDA PROTEIN APE0766.
 GN APE0766.
 OS Aeropyrum pernix.
 CC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 CC Aeropyrum.
 NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KI;
 RX MEDLINE=99310339; PubMed=10382960;
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Aikawa A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.,
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix KI."
 RL DNA Res. 6:83-101(1999).

DR EMBL: AP000060; BAA79744.1;
 DR InterPro: IPR001454; Hydrolase.
 DR InterPro: IPR000150; Hypothet_cof.
 DR Pfam: PF00702; Hydrolase; 1.
 DR PROSITE: PS01229; COF_2; UNKNOWN_1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 239 AA; 24758 MW; EA30A3CDF86B3DE CRC64;

Query Match 85.7%; Score 30; DH 1; Length 239;
 Hest Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKXWR 6
 DB 105 FKPSWQ 110
 1:1 1;
 105 FKPSWQ 110

RESULT 10
 QYD9R1 PRELIMINARY; PRT; 440 AA.
 AC QYD9R1;
 DT 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DE ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
 DE CLONE:1700030N20, FULL INSERT SEQUENCE.
 CN 5830420C20R1K.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TESTIS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Hatalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauble F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Hoffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Hult C., Fletcher C., Carninci P., Fujita M., Gariboldi M.,
 RA Guslinich S., Hill D., Holmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AK006565; HAB24655.1;
 DR MGI: 1923275; 5830420C20R1K.
 SQ SEQUENCE 440 AA; 49566 MW; 5B5DA87E19879644 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 440;
 Hest Local Similarity 66.7%; Pred. No. 3.5e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 FKXWR 6
 DB 39 FRPRMR 44
 1:1 11
 39 FRPRMR 44

RESULT 11
 QY1566 PRELIMINARY; PRT; 606 AA.
 ID QY1566

AC 031566;
 DT 01-JAN-1998 (TReMBLrel. 05, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
 DE YFIX PROTEIN.
 CN YFIX.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=9804033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borliss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Broutillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Chim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina M., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portwiliik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RI Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 299108; CAB12672.1; -;
 KW Complete proteome.
 SQ SEQUENCE 606 AA; 68677 MW; 4B58DD58922ED0CA CRC64;

Query Match 85.7%; Score 30; DB 2; Length 606;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 DB 570 YKPEWR 575
 :|||

RESULT 12
 O52961
 ID O52961 PRELIMINARY: PRT: 610 AA.
 AC O52961;
 DT 01-JUN-1998 (TReMBLrel. 06, Created)
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
 DT 01-NOV-1998 (TReMBLrel. 08, Last annotation update)
 DE YFIX.
 OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97101647; PubMed=8946165;
 RA Yamamoto H., Uchiyama S., Sekiguchi J.,
 RI "Cloning and sequencing of a 27.8-kb nucleotide sequence of the 79
 PT degrees-81 degrees region of the Bacillus subtilis genome containing
 RT the sspE locus.";
 RL DNA Res. 3:257-262(1996).
 DR EMBL; D85082; BAA24464.1; -;
 SQ SEQUENCE 610 AA; 69089 MW; 34094DD5CA72FF66 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 610;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 DB 574 YKPEWR 579
 :|||

RESULT 13
 O9NA80 PRELIMINARY: PRT: 684 AA.
 ID O9NA80;
 AC O9NA80;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
 DE Y57A10A.20 PROTEIN.
 GN Y57A10A.20.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smyle R.;
 RI Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RL Submits.";
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RI investigating biology.";
 RL Science 282:2012-2018(1998).
 DR EMBL; AL117195; CAB55023.1; -;
 SQ SEQUENCE 684 AA; 79659 MW; F0E4FAF28EAF687 CRC64;

Query Match 85.7%; Score 30; DB 5; Length 684;
 Best Local Similarity 66.7%; Pred. No. 5.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 5
 DB 672 FKPKWK 677
 :|||

RESULT 14
 Q12361
 ID Q12361 PRELIMINARY: PRT: 961 AA.
 AC Q12361;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 110.7 KDA PROTEIN YDL035C.
 GN GPI OR D2749 OR YDL035C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ALPHA S288C;
RA Paulin L., Saron A.M.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBDJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Paulin L., Saron A.M., Laamanen P.;
RI Submitted (JUL-1996) to the EMBL/GenBank/DBDJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBDJ databases.
DR EMBL; Z71781; CAA96454.1; -;
DR EMBL; Z74083; CAA98593.1; -;
DK SCD; S0002193; GPR1.
KW Hypothetical protein.
SQ SEQUENCE 961 AA: 110708 MW: 9899D857872A4209 CRC64;

Query Match 85.7%; Score 30; DH 3; Length 961;
Best Local Similarity 66.7%; Pred. No. 7.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
DB 156 FKNWK 161

RESULT 15

Q9LJQ1 Q9LJQ1 PRELIMINARY; PRT: 1232 AA.
AC Q9LJQ1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE G01A055299.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBDJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety pl,
RT TAC and BAC clones.";
HL DNA RGS. 7:217-221(2000).
CC -!- SIMILARITY: CONTAINS A RING-TYPE ZINC FINGER.
DR EMBL; AP000414; HA01179.1; -;
DR InterPro; IPR000194; ATPase_alpha_beta.
DR InterPro; IPR000345; CytC_name_bind.
DR InterPro; IPR001841; ZnF_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00184; KINC; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Zinc-finger.
SQ SEQUENCE 1232 AA: 139223 MW: D3770C4A9D699207 CRC64;

Query Match 85.7%; Score 30; DH 10; Length 1232;
Best Local Similarity 66.7%; Pred. No. 9.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
DB 992 FKPCWK 907

Search completed: February 27, 2002, 11:50:09
Job time: 988 sec

Thu Feb 28 11:23:42 2002

us-09-446-109a-8.rspt

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:04 ; Search time 132.19 Seconds
(without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109A-8
Perfect score: 35
Sequence: 1 FKXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues
Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2.6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2.6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2.6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2.6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2.6/prodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2.6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length	ID	Description
1	33	94.3	6	1 US-08-215-137-13	Sequence 13, Appl
2	28	80.0	13	4 US-08-915-314-26	Sequence 26, Appl
3	28	80.0	16	1 US-08-079-051-2	Sequence 2, Appl
4	28	80.0	16	5 PCT-US94-06994-2	Sequence 2, Appl
5	28	80.0	115	3 US-08-513-974B-351	Sequence 351, App
6	28	80.0	243	2 US-08-829-110-3	Sequence 3, Appl
7	28	80.0	304	1 US-08-118-270-35	Sequence 35, Appl
8	28	80.0	304	5 PCT-US93-08528-35	Sequence 35, Appl
9	28	80.0	350	2 US-08-458-970A-9	Sequence 9, Appl
10	28	80.0	463	3 US-09-082-310-1	Sequence 1, Appl
11	28	80.0	482	2 US-08-876-874-2	Sequence 2, Appl
12	28	80.0	496	3 US-08-906-769-113	Sequence 113, App
13	28	80.0	496	3 US-08-906-616-113	Sequence 113, App
14	28	80.0	496	4 US-08-817-795-113	Sequence 113, App
15	28	80.0	496	4 US-08-639-075A-113	Sequence 113, App
16	28	80.0	496	4 US-09-012-431-113	Sequence 113, App
17	28	80.0	496	4 US-09-012-692-113	Sequence 113, App
18	28	80.0	496	4 US-08-906-613-113	Sequence 113, App
19	28	80.0	496	5 PCT-US95-14442A-113	Sequence 113, App
20	28	80.0	592	3 US-08-991-813-2	Sequence 2, Appl
21	27	77.1	157	4 US-09-461-474-14	Sequence 14, Appl
22	27	77.1	355	2 US-08-666-367B-6	Sequence 6, Appl
23	27	77.1	355	4 US-09-143-438-6	Sequence 6, Appl
24	27	77.1	448	4 US-09-461-474-8	Sequence 8, Appl
25	27	77.1	474	4 US-09-461-474-10	Sequence 10, Appl
26	27	77.1	527	3 US-08-907-229-2	Sequence 2, Appl
27	27	77.1	566	2 US-08-666-467B-5	Sequence 5, Appl

24 27 77.1 566 4 US-09-143-438-5 Sequence 5, Appl
25 27 77.1 1040 2 US-08-254-989-2 Sequence 2, Appl
30 27 77.1 1536 4 US-09-413-814-10 Sequence 10, Appl
31 27 77.1 2232 4 US-09-091-219-25 Sequence 25, Appl
32 27 77.1 2247 4 US-09-091-219-2 Sequence 2, Appl
33 26 74.3 10 2 US-08-704-655-20 Sequence 20, Appl
34 26 74.3 37 1 US-08-665-543B-4 Sequence 4, Appl
35 26 74.3 42 2 US-08-766-858A-27 Sequence 27, Appl
36 26 74.3 209 4 US-09-164-193-8 Sequence 8, Appl
37 26 74.3 273 2 US-08-997-080-75 Sequence 75, Appl
38 26 74.3 273 2 US-08-997-362-75 Sequence 75, Appl
39 26 74.3 273 3 US-08-873-970-75 Sequence 75, Appl
40 26 74.3 273 4 US-09-095-855-75 Sequence 75, Appl
41 26 74.3 336 4 US-09-334-601-13 Sequence 13, Appl
42 26 74.3 370 2 US-08-997-080-194 Sequence 194, App
43 26 74.3 370 2 US-08-997-362-194 Sequence 194, App
44 26 74.3 370 4 US-09-095-855-194 Sequence 194, App
45 26 74.3 417 1 US-08-351-981-6 Sequence 6, Appl

ALIGNMENTS

RESULT: 1
US-08-215-137-13
; Sequence 13, Application US/08215137
; Patent No. 5614370
; GENERAL INFORMATION:
; APPLICANT: Konleatlis, Zenon
; APPLICANT: Siciliano, Salvatore J
; APPLICANT: Springer, Martin S
; TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN C5A ANTAGONISTS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: 126 Lincoln Ave., P.O. Box 2000
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: United States of America
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/215,137
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Benesh, Gerard H
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: 19108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 594-3901
; TELEFAX: (908) 594-4720
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MUTATION TYPE: peptide
; SYNTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1
; OTHER INFORMATION: /label= BII
; OTHER INFORMATION: /note= either the natural phenylalanine amino
; OTHER INFORMATION: terminus or the Holton-Hunter modified peptide

OTHER INFORMATION: having the group 3-(p-hydroxyphenyl)propionyl group

FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label- dCha
OTHER INFORMATION: /note- "D-cyclohexylalanine"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /label- dArg
OTHER INFORMATION: /note- "D-arginine"
US-08-215-137-13

Query Match 94.3%; Score 33; DB 1; Length 6;
Best Local Similarity 100.0%; Pred. No. 1.6e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 FKPXWR 6
Db 1 FKPXWR 6

RESULT 2
US-08-915-314-26
Sequence 26, Application US/08915314
Patent No. 6180604
GENERAL INFORMATION:
APPLICANT: Fraser, Janet R.
APPLICANT: West, Michael H.P.
APPLICANT: Krieger, Timothy J.
APPLICANT: Taylor, Robert
APPLICANT: Erfile, Douglas
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
TITLE OF INVENTION: INFECTIONS USING ANALOGUES OF INDOLICIDIN
NUMBER OF SEQUENCES: 90
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,314
FILING DATE: 20-AUG-1997
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: No. 6180604tenburg ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 660081.405
TELEPHONE INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 26:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
US-08-915-314-26

Query Match 80.0%; Score 28; DB 4; Length 13;
Best Local Similarity 66.7%; Pred. No. 8.3;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXWR 6

Db 4 WKPXWR 9

RESULT 3
US-08-079-051-2
Sequence 2, Application US/08079051
Patent No. 5480974
GENERAL INFORMATION:
APPLICANT: MORGAN, EDWARD L.
APPLICANT: EMBER, JULIA A.
APPLICANT: HUGLI, TONY E.
TITLE OF INVENTION: ANTIBODIES TO HUMAN C5a RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/079,051
FILING DATE: 18-JUN-1993
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Wetherell, Jr., Ph.D., John W.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: PD-2790
TELEPHONE INFORMATION:
TELEPHONE: (619) 455-5110
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: C5aR(133-148)
FEATURE:

NAME/KEY: Peptide
LOCATION: 1..16

US-08-079-051-2

Query Match 80.0%; Score 28; DB 1; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXW 5
Db 7 FKPIW 11

RESULT 4
PCT-US94-06994-2
Sequence 2, Application PC/TUS9406994
GENERAL INFORMATION:
APPLICANT: The Scripps Research Institute
TITLE OF INVENTION: ANTIBODIES TO HUMAN C5a RECEPTOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Spensley Horn Jubas & Lubitz
STREET: 1880 Century Park East, Suite 500
CITY: Los Angeles
STATE: California

COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06994
FILING DATE: 20-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Tumarkin, Lisa A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: FD-2790
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: C5aR(133-148)
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
PCT-US94-06994-2

Query Match 80.0%; Score 28; DB 5; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXW 5
DB 7 FRPIW 11

RESULT 5
US-08-513-974B-351
Sequence 351, Application US/08513974B
Patent No. 6114139
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Ohtaki, Tetsuya
APPLICANT: Fukusumi, Shoji
APPLICANT: Ogi, Kazuhiro
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
TITLE OF INVENTION: PRODUCTION, AND USE THEREOF
NUMBER OF SEQUENCES: 380
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/513,974B
FILING DATE: 14-SEP-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-093989
FILING DATE: 19-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-270017
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236357
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236356
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189274
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189273
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45753
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 351:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-513-974B-351

Query Match 80.0%; Score 28; DB 3; Length 115;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXW 5
DB 111 1
; FRPIW 8

RESULT 6
US-03 429-110-3
Sequence 3, Application US/08829110
Patent No. 5842390
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL REGULATORS OF G-PROTEIN
TITLE OF INVENTION: SIGNALING
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 1174 Porter Drive
CITY: Palo Alto
STATE: CA


```

: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/29,110
: FILING DATE: Filed Herewith
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PP-0259 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 243 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: TUMNOT02
: CLONING: 343504
: US-08-829-110-3

Query Match 80.0%; Score 28; DB 2; Length 243;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KPXR 6
Db 71 KPAWR 75

RESULT 7
US-08-118-270-35
: Sequence 35, Application US/08/118270
: Patent No. 5508384
: GENERAL INFORMATION:
: APPLICANT: Murphy, Randall H.
: TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
: NUMBER OF SEQUENCES: 348
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/118,270
: FILING DATE: 09-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/943,236
: FILING DATE: 10-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Townsend, Kevin G.
: REGISTRATION NUMBER: 34,033
: REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 35:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 304 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: PCT-US93-08528-35

Query Match 80.0%; Score 28; DB 5; Length 304;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXW 5
Db 103 FKPIW 107

REFERENCE/DOCKET NUMBER: MURPHY-2A
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 35:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 304 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-118-270-35

Query Match 80.0%; Score 28; DB 1; Length 304;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXW 5
Db 103 FKPIW 107

RESULT 8
PCT-US93-08528-35
: Sequence 35, Application PC/TUS93/08528
: GENERAL INFORMATION:
: APPLICANT: New York University
: TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
: NUMBER OF SEQUENCES: 348
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W., Suite 300
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/08528
: FILING DATE: 09-SEP-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/943,236
: FILING DATE: 10-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Townsend, Kevin G.
: REGISTRATION NUMBER: 34,033
: REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: TELEX: 248633
: INFORMATION FOR SEQ ID NO: 35:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 304 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: PCT-US93-08528-35

Query Match 80.0%; Score 28; DB 5; Length 304;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXW 5
```

Db 103 FKPIW 107

RESULT 9
US-08-458-970A-9
; Sequence 9, Application US/08458970A
; Patent No. 5861272
; GENERAL INFORMATION:
; APPLICANT: I.I. ET AL.
; TITLE OF INVENTION: C5a Receptor
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 HECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,970A
; FILING DATE: June 2, 1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09234
; FILING DATE: 16 AUG 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-353
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 350 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-458-970A-9

Query Match 80.0%; Score 28; DB 2; Length 350;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPYW 5
Db 139 FKPIW 143

RESULT 10
US-08-310-1
; Sequence 1, Application US/09082310
; Patent No. 6096526
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: HUMAN NUCLEIC ACID METHYLASES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE

CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/082,310
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0520 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BKSTNOT07
; CLONE: 2124957
US-09-082-310-1

Query Match 80.0%; Score 28; DB 3; Length 463;
Best Local Similarity 80.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KPXWR 6
Db 293 KPSWR 297

RESULT 11
US-08-876-874-2
; Sequence 2, Application US/08876874
; Patent No. 5942405
; GENERAL INFORMATION:
; APPLICANT: Amos, Robert
; APPLICANT: Bergsma, Derk
; APPLICANT: Poley, James
; APPLICANT: Kumar, Chandrika
; TITLE OF INVENTION: THERAPEUTIC AND SCREENING
; TITLE OF INVENTION: METHODS USING C3A RECEPTOR AND C3A
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RATNER & PRESTIA
; STREET: P.O. BOX 980
; CITY: VALLEY FORGE
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: EasySEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,874
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/019,627
; FILING DATE: 16-JUN-1997

ATTORNEY/AGENT INFORMATION:

NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: P50501
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 482 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-876-874-2

Query Match 80.0%; Score 28; DB 2; Length 482;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPXW 5

Db 125 FKPIW 129

RESULT 12

US-08-906-769-113
Sequence 113, Application US/08906769
Patent No. 6077687

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary

TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/08/906.769

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/639,075

FILING DATE: 24-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-25-C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 113:

SEQUENCE CHARACTERISTICS:

LENGTH: 496 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-906-769-113

Query Match 80.0%; Score 28; DB 3; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPXW 5

Db 358 FKPKW 362

RESULT 13

US-08-906-616-113

Sequence 113, Application US/08906616

Patent No. 6121035

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.

APPLICANT: Rushlow, Keith E.

APPLICANT: Wu Hunter, Shirley

APPLICANT: Frank, Glenn R.

APPLICANT: Stiegler, Gary

APPLICANT: Gaines, Patrick J.

APPLICANT: Silver, Gary

TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF

NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sheridan Ross P.C.

STREET: 1700 Lincoln Street, Suite 3500

CITY: Denver

STATE: Colorado

COUNTRY: USA

ZIP: 80203

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION NUMBER: US/08/906.616

FILING DATE: 05-AUG-1997

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Connell, Gary J.

REGISTRATION NUMBER: 32,020

REFERENCE/DOCKET NUMBER: 2618-25-C2-3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (303) 863-9700

TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 113:

SEQUENCE CHARACTERISTICS:

LENGTH: 496 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-906-616-113

Query Match 80.0%; Score 28; DB 3; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 PKPXW 5

Db 358 FKPKW 362

RESULT 14

US-08-877-795-113

Sequence 113, Application US/08817795

Patent No. 6139840

GENERAL INFORMATION:

APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Frank, Glenn R.
APPLICANT: Heath, Andrew W.
APPLICANT: Yamaka, Mico Yamanaka
APPLICANT: Arfsten, Ann
APPLICANT: Dale, Beverly
APPLICANT: Sulegler, Gary
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES, AND USES THEREOF
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/817,795
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: PCT/US95/14442
APPLICATION NUMBER: 32,020
ATTORNEY/AGENT INFORMATION:
NAME: Gary J. Connell
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-817-795-113

Query Match 80.0%; Score 28; DB 4; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKW 5
DB 358 FKPKW 362

RESULT 15
US-08-639-075A-113
Sequence 113, Application US/08639075A
Patent No. 6150125
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-639-075A-113

Query Match 80.0%; Score 28; DB 4; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKW 5
DB 358 FKPKW 362

Search completed: February 27, 2002, 11:36:04
Job time: 143 sec

27 USC-1993, 90US-0634641.

XX (ABBO) ABBOTT LABORATORIES.
 XX Kawai M, Luly JR, Or YS, Wagner R, Wiedeman PE;
 XX WPI: 1992-268383/32.
 XX New hexa- and hepta-peptide(s) are anaphylatoxin antagonists and
 PT agonists - for treating inflammatory and immunodeficiency
 PT diseases, cancers and severe infections
 XX Claim 11: Page 153: 160pp: English.
 XX The peptide is a specifically claimed example of a group of highly
 CC generic hexa- and heptapeptides which are (a) anaphylatoxin
 CC antagonists useful for treating asthma, other allergies,
 CC inflammations, autoimmune diseases, serum sickness, gout, bullous
 CC skin diseases, psoriasis, ARDS, endotoxin shock, hepatic cirrhosis,
 CC pancreatitis, inflammatory bowel disease, burns, sepsis, myocardial
 CC infarction, chronic hepatitis, transplant rejection, or ischaemic
 CC heart or brain damage; or (b) anaphylatoxin agonists useful for
 CC stimulating inflammatory and immune responses, e.g. in the treatment
 CC of cancer, immunodeficiency diseases and severe infections.
 XX Sequence 6 AA:
 SQ

Query Match 94.3%; Score 33; DB 13; Length 6;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRPXWR 6
 Db 1 fkpwxr 6

RESULT 2
 AAR89737
 ID AAR89737 standard; peptide: 6 AA.
 AC AAR89737;
 XX 03-MAY-1996 (first entry)
 DT C5a peptide analogue, C089, used to identify C5a (ant)agonists.
 DE C5a: complement; agonist; antagonist; peptide; binding assay;
 XX identification; inflammation; pain reduction; respiratory disorder;
 KW cardiovascular; hypertension; ARDS; anaphylaxis; arthritis; asthma;
 KW competitive.
 XX Synthetic.
 OS Key Location/Qualifiers
 XX Misc-difference 1
 FT /label= phe, OTHER
 FT /note= "OTHER = the Bolton-Hunter modified peptide
 FT having the 3-(p-hydroxyphenyl)-propionyl
 FT group".
 FT Misc-difference 4
 FT /label= OTHER
 FT /note= "D-cyclohexylalanine"
 FT Modified-site 6
 FT /label= OTHER
 FT /note= "Cyclohexylalanine"
 XX WO9525957-A1.
 XX 28-SEP-1995.
 XX 14-MAR-1995: 95WO-0503209.
 XX 18-MAR-1994: 94US-0215137.

XX (MERI) MERCK & CO INC.
 XX Konteatis 2, Siciliano SJ, Springer MS;
 XX WPI: 1995-344718/44.
 XX C5a receptor binding assays - used for identifying cpds.having C5a
 PT antagonist, agonist or partial agonist activity
 XX Claim 15: Page 47: 65pp: English.
 XX C5a C-terminal peptide analogues AAR89734-44 and AAR90033-35 may be used
 CC in a new method for identifying C5a agonists/antagonists AAR89734
 CC represents the generic formula of this newly defined class of
 CC labelled peptides. C5a antagonists identified using this method are
 CC useful in the treatment of a wide variety of C5a-mediated diseases
 CC such as acute respiratory distress syndrome (ARDS), anaphylactic
 CC shock, psoriasis, osteoarthritis, rheumatoid arthritis, asthma,
 CC cardiovascular disorders, and metastatic spread of cancerous tumours.
 XX Sequence 6 AA:
 SQ

Query Match 94.3%; Score 33; DB 16; Length 6;
 Best Local Similarity 100.0%; Pred. No. 4.3e+05;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FRPXWR 6
 Db 1 fkpwxr 6

RESULT 3
 AAW74031
 ID AAW74031 standard; peptide: 39 AA.
 XX AAW74031;
 XX 04-MAY-1999 (first entry)
 DT Human D2H binding protein DAB3.
 DE Gastro-intestinal transport receptor; binding protein; hst; HPT1;
 XX D2H; hPPT1; human; GI tract receptor; sucrose-isomaltase complex;
 KW intestinal peptide-associated transporter; hypertension; diabetes;
 KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
 XX therapeutic agent delivery; therapy.
 XX Homo sapiens.
 OS WO951325-A2.
 XX 19-NOV-1998.
 XX 15-MAY-1998: 98WO-US10088.
 XX 15-MAY-1997: 97US-0046595.
 XX (CYTO-) CYTOGEN CORP.
 PA (ELAN-) ELAN CORP PLC.
 XX Alvarez VL, Bellinka BA, Cagney GM, Carter JM, Lambkin IJ;
 PI Omahony DJ, Patterson CA, Singleton J;
 XX WPI: 1999-009568/01.
 XX New proteins that bind specifically to receptors in the
 PT gastro-intestinal tract and related nucleic acid - chimaeras and
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide
 XX Claim 2: Page 54: 294pp: English.

XX This sequence represents a peptide that specifically binds to the human
 CC D2H protein. The invention relates to purified proteins (I) that bind
 CC specifically to at least one of the gastro-intestinal (GI) tract
 CC receptors human intestinal peptide-associated transporter (HPT1),
 CC hPPT1, D2H and human sucrose-isomaltase complex (hst). (I) provide
 CC active transport of therapeutic agents through human and animal GI tissue
 CC (into the blood) for in vivo delivery, particularly for treatment or
 CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
 CC cancer, migraine, or angina pectoris. Specifically they are used to
 CC deliver insulin or leuprolide, but many other suitable therapeutic agents
 CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
 CC and antigens. (I) may also provide targeting to the GI tract. Other uses
 CC of (I) are: (i) to determine the level of specified receptors in a sample
 CC (in a binding assay); and (ii) to screen for molecules that bind (I).
 CC Immunogenic analogues or derivatives of (I) are used to raise antibodies
 CC and in immunoassays. The antibodies are used to locate, detect and
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
 CC etc., also for peptide purification and immobilisation.
 XX Sequence 39 AA:

Query Match 85.7%; Score 30; DB 20; Length 39;
 Best Local Similarity 66.7%; Pred. No. 18;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
 DB 15 frpgwr 20

RESULT 4

AA801208
 ID AAB01208 standard; Protein: 434 AA.

XX AA801208;

XX 12-DEC-2000 (first entry)

XX Corn putative lecithin:cholesterol acyltransferase #3.

XX Corn: lecithin:cholesterol acyltransferase; phytosterol;
 KW phosphatidylcholine-sterol O-acyltransferase; heat shock; cold shock.

XX Zea mays.

XX WO200032791-A2.

XX 08-JUN-2000.

XX 02-DEC-1999; 99WO-US28586.

XX 03-DEC-1998; 98US-0110782.

XX (DUPO) DU PONT DE NEMOURS & CO E. I.

XX Cahoon RE, Kinney AJ, Sakai II, Shen JB, Butler KH, Saylor JJ;

XX WPI; 2000-412337/35.

XX N-PSDB; AAA49203.

XX Polynucleotide encoding plant lecithin:cholesterol acyltransferase
 PT enzyme useful for producing transgenic plants and for producing
 PT antibodies specific to which is useful for screening cDNA expression
 PT libraries.

XX Claim 10; Page 40-41; 49pp; English.

XX The present sequence is a putative protein sequence of a corn

CC lecithin:cholesterol acyltransferase (also known as

CC phosphatidylcholine-sterol O-acyltransferase). This enzyme is found
 CC associated with high-density lipoproteins and Apolipoprotein-AI and -D.

CC The gene and protein can be used to produce transgenic plants which have
 CC increased lipid metabolism and membrane fluidity, and therefore increased
 CC resistance to heat and/or cold shock, to alter the content of phytosterol
 CC or lecithin in grains and to identify potential herbicides.

XX Sequence 434 AA;

Query Match 85.7%; Score 30; DB 21; Length 434;

Best Local Similarity 66.7%; Pred. No. 2.1e+02;

Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
 DB 271 frpmwr 276

RESULT 5

AAG98875
 ID AAG98875 standard; Protein: 462 AA.

XX AC

XX AA98875;

XX 26-SEP-2001 (first entry)

XX 2. coli growth and proliferation related protein sequence SEQ ID NO:345.

XX Escherichia coli; growth; proliferation; microbial; antimicrobial;

XX bacterial infection; microorganism.

XX Escherichia coli.

XX P0230134810-A2.

XX 17-MAY-2001.

XX 09 NOV-2000; 2000WO-US30950.

XX 09-NOV-1999. 99US-0164415.

XX (EITF-) ELITRA PHARM INC.

XX Romsyth RA, Ohlsen K, Zyskind J;

XX WPI; 2001-335933/35.

XX W-PSDB; AAH84546.

XX Acyl nucleic acids that inhibit Escherichia coli proliferation, useful
 PT for screening for homologous genes and for designing expression vectors

XX Claim 19; Page 421-422; 522pp; English.

XX AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
 CC related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
 CC growth and proliferation related proteins given in AAG99078 and AAG98830
 CC to AAG98999. (I) can be used as potential targets for the generation of
 CC new antimicrobial agents, and for identification of compounds which
 CC interact with the gene products of (I). In addition the expression of
 CC (I) and the purification of the proteins, the purified proteins can be
 CC used to generate reagents and screen small molecule libraries or other
 CC candidate compound libraries for compounds that can be further developed
 CC to yield novel antimicrobial compounds. In addition, nucleic acid probes
 CC complementary to (I) that are specific for particular species of
 CC microorganisms can be used to identify particular microorganism species
 CC in clinical specimens, therefore, providing a rapid and dependable
 CC method by which to identify the causative agents of a bacterial
 CC infection. Also, antibodies generated against proteins translated from
 CC mRNA transcribed from proliferation-required sequences can also be used
 CC to screen for specific microorganisms that produce such proteins in a
 CC species-specific manner. AAH84371 and AAH84670 represent sequencing
 CC primers used in the isolation of E. coli growth and proliferation
 CC related sequence, which are used in an example from the present


```
CC invention.
XX
SQ Sequence 462 AA;

Query Match      85.7%; Score 30; DB 22; Length 462;
Best Local Similarity 66.7%; Pred. No. 2.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPxWR 6
   :|||
Db 398 ykplwr 403

RESULT 6
AAY58359
ID AAY58359 standard; Protein: 700 AA.
XX
AC AAY58359;
XX
DT 27-MAR-2000 (first entry)
XX
DE Yeast G protein-coupled receptor GPR1 mutant d694-954.
XX
KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
KW pseudohyphal form; signalling pathway; antifungal; mutant; mutein.
XX
OS Saccharomyces cerevisiae.
OS Synthetic.
XX
PN WO9663094-A2.
XX
PD 09-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11838.
XX
PR 01-JUN-1998; 98US-0088311.
XX
PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
PI Hirsch JP, Xue Y;
XX
WPI: 2000-086980/07.
XX
Novel protein, gene, antibody and recombinant cell useful for
identifying modulators and inhibitors useful as anti-fungals -
Example 6.2.5; Page -: 63pp; English.
XX
The invention relates to a novel yeast G protein-coupled receptor,
GPR1 (AAY58355). GPR1 is predicted to have seven transmembrane domains,
which is characteristic of G protein-coupled receptors, and also
contains a very large third cytoplasmic loop and a large cytoplasmic
tail. The third cytoplasmic loop contains two short basic sequences; one
sequence (KRKAQIG) is present at the N-terminal end of the loop and the
other sequence (KKRAQIQ) is at the C-terminal end of the loop. The
third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
believed to provide the upstream signal that activates GPA2, and plays
an essential role in inducing the switch from non-filamentous to
filamentous growth in yeast. Modulators of GPR1 function, especially
inhibitors, are useful for prevention or treatment of a disease or
disorder involving a fungal infection. The inhibitor may be an anti-GPR1
antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
inserted heterologous sequence inactivates the biological activity of
GPR1. Compounds which modulate GPR1 gene expression and/or gene product
activity may be used to inhibit the conversion of a fungus from a
non-filamentous form to a filamentous form. The present sequence
represents a GPR mutant, d277-284, used in an exemplification of
the invention. This sequence is lacking a portion (designated the
C-terminal tail, relative to the native GPR1.
Note: This sequence is not shown in the specification, but is derived
from the yeast GPR1 amino acid sequence given in figure 1A.
XX
```

```
SQ Sequence 700 AA;

Query Match      85.7%; Score 30; DB 21; Length 700;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPxWR 6
   :|||
Db 156 fkpwnk 161

RESULT 7
AAY58360
ID AAY58360 standard; Protein: 847 AA.
XX
AC AAY58360;
XX
DT 27-MAR-2000 (first entry)
XX
DE Yeast G protein-coupled receptor GPR1 mutant d841-954.
XX
KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
KW pseudohyphal form; signalling pathway; antifungal; mutant; mutein.
XX
OS Saccharomyces cerevisiae.
OS Synthetic.
XX
PN WO9663094-A2.
XX
PD 09-DEC-1999.
XX
PF 23-MAY-1999; 99WO-US11838.
XX
PR 01-JUN-1998; 98US-0088311.
XX
PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
PI Hirsch JP, Xue Y;
XX
WPI: 2000-086980/07.
XX
Novel protein, gene, antibody and recombinant cell useful for
identifying modulators and inhibitors useful as anti-fungals -
Example 6.2.5; Page -: 63pp; English.
XX
The invention relates to a novel yeast G protein-coupled receptor,
GPR1 (AAY58355). GPR1 is predicted to have seven transmembrane domains,
which is characteristic of G protein-coupled receptors, and also
contains a very large third cytoplasmic loop and a large cytoplasmic
tail. The third cytoplasmic loop contains two short basic sequences; one
sequence (KRKAQIG) is present at the N-terminal end of the loop and the
other sequence (KKRAQIQ) is at the C-terminal end of the loop. The
third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
believed to provide the upstream signal that activates GPA2, and plays
an essential role in inducing the switch from non-filamentous to
filamentous growth in yeast. Modulators of GPR1 function, especially
inhibitors, are useful for prevention or treatment of a disease or
disorder involving a fungal infection. The inhibitor may be an anti-GPR1
antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
inserted heterologous sequence inactivates the biological activity of
GPR1. Compounds which modulate GPR1 gene expression and/or gene product
activity may be used to inhibit the conversion of a fungus from a
non-filamentous form to a filamentous form. The present sequence
represents a GPR mutant, d277-284, used in an exemplification of
the invention. This sequence is lacking a portion (designated the
C-terminal tail, relative to the native GPR1.
Note: This sequence is not shown in the specification, but is derived
from the yeast GPR1 amino acid sequence given in figure 1A.
XX
```

```
Query Match      85.7%; Score 30; DB 21; Length 847;
Best Local Similarity 66.7%; Pred. No. 4.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPNXR 6
DB 156 fkpnrk 161

RESULT 8
ID AAY58358 Standard; Protein: 864 AA.
XX
AC AAY58358;
XX
DT 27-MAR-2000 (first entry)
XX
DE Yeast G protein-coupled receptor GPR1 mutant d490-586.
XX
KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
KW pseudohyphal form; signalling pathway; antifungal; mutant; mutuin.
XX
OS Saccharomyces cerevisiae.
OS Synthetic.
XX
PW W09963094-A2.
XX
PN
PD 09-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11838.
XX
PR 01-JUN-1998; 98US-0088311.
XX
PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
PI Hirsch JP, Xue Y;
XX
DR WPI; 2000-086980/07.
XX
PT Novel protein, gene, antibody and recombinant cell useful for
PT identifying modulators and inhibitors useful as anti-fungals -
XX
PS Example 6.2.5; Page -: 63pp; English.
XX
CC The invention relates to a novel yeast G protein-coupled receptor,
CC GPR1 (AAY58358). GPR1 is predicted to have seven transmembrane domains,
CC which is characteristic of G protein-coupled receptors, and also
CC contains a very large third cytoplasmic loop and a large cytoplasmic
CC tail. The third cytoplasmic loop contains two short basic sequences; one
CC other sequence (KRKAQIG) is present at the N-terminal end of the loop and the
CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
CC believed to provide the upstream signal that activates GPA2, and plays
CC an essential role in inducing the switch from non-filamentous to
CC filamentous growth in yeast. Modulators of GPR1 function, especially
CC inhibitors, are useful for prevention or treatment of a disease or
CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
CC inserted heterologous sequence inactivates the biological activity of
CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
CC activity may be used to inhibit the conversion of a fungus from a
CC non-filamentous form to a filamentous form. The present sequence
CC represents a GPR mutant, d277-284, used in an exemplification of
CC the invention. This sequence is lacking the asparagine-rich region
CC of the third cytoplasmic loop, relative to the native GPR1.
CC Note: This sequence is not shown in the specification, but is derived
CC from the yeast GPR1 amino acid sequence given in figure 1A.
XX
SQ Sequence 864 AA;
```

```
Query Match      85.7%; Score 30; DB 21; Length 953;
Best Local Similarity 66.7%; Pred. No. 4.7e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPNXR 6
DB 156 fkpnrk 161

RESULT 9
ID AAY58356 Standard; Protein: 953 AA.
XX
AC AAY58356;
XX
DT 27-MAR-2000 (first entry)
XX
DE Yeast G protein-coupled receptor GPR1 mutant d277-284.
XX
KW Yeast; G protein-coupled receptor; GPR1; filamentous growth;
KW pseudohyphal form; signalling pathway; antifungal; mutant; mutuin.
XX
OS Saccharomyces cerevisiae.
OS Synthetic.
XX
PW W09963094-A2.
XX
PN
PD 09-DEC-1999.
XX
PF 28-MAY-1999; 99WO-US11838.
XX
PR 01-JUN-1998; 98US-0088311.
XX
PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
PI Hirsch JP, Xue Y;
XX
DR WPI; 2000-086980/07.
XX
PT Novel protein, gene, antibody and recombinant cell useful for
PT identifying modulators and inhibitors useful as anti-fungals -
XX
PS Example 6.2.5; Page -: 63pp; English.
XX
CC The invention relates to a novel yeast G protein-coupled receptor,
CC GPR1 (AAY58356). GPR1 is predicted to have seven transmembrane domains,
CC which is characteristic of G protein-coupled receptors, and also
CC contains a very large third cytoplasmic loop and a large cytoplasmic
CC tail. The third cytoplasmic loop contains two short basic sequences; one
CC other sequence (KRKAQIG) is present at the N-terminal end of the loop and the
CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
CC believed to provide the upstream signal that activates GPA2, and plays
CC an essential role in inducing the switch from non-filamentous to
CC filamentous growth in yeast. Modulators of GPR1 function, especially
CC inhibitors, are useful for prevention or treatment of a disease or
CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
CC inserted heterologous sequence inactivates the biological activity of
CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
CC activity may be used to inhibit the conversion of a fungus from a
CC non-filamentous form to a filamentous form. The present sequence
CC represents a GPR mutant, d277-284, used in an exemplification of
CC the invention. This sequence is lacking the first basic motif (KRKAQIG)
CC of the third cytoplasmic loop, relative to the native GPR1.
CC Note: This sequence is not shown in the specification, but is derived
CC from the yeast GPR1 amino acid sequence given in figure 1A.
XX
SQ Sequence 953 AA;
```

Matches	4:	Conservative	1:	Mismatches	1:	Indels	0:	Gaps	0:
QY	1	FKPXWR 6							
Db	156	fkpnwk 161							
RESULT 10									
ID	AAV58357	standard; Protein: 953 AA.							
XX	AC	AAV58357;							
XX	DT	27-MAR-2000 (first entry)							
XX	DE	Yeast G protein-coupled receptor GPR1 mutant d610-617.							
XX	KW	Yeast: G protein-coupled receptor; GPR1; filamentous growth; pseudohyphal form; signalling pathway; antifungal; mutant; mutein.							
XX	KW	Saccharomyces cerevisiae.							
XX	OS	Synthetic.							
XX	PN	WO9963094-A2.							
XX	PD	09-DEC-1999.							
XX	XX	28-MAY-1999; 99WO-US11838.							
XX	XX	01-JUN-1998; 98US-0088311.							
XX	PA	(MOUN) MOUNT SINAI SCHOOL MEDICINE.							
XX	PI	Hirsch JP, Xue Y;							
XX	DR	WPI: 2000-086980/07.							
XX	PT	Novel protein, gene, antibody and recombinant cell useful for identifying modulators and inhibitors useful as anti-fungals -							
XX	PS	Example 6.2.5; Page -: 63pp; English.							
XX	CC	The invention relates to a novel yeast G protein-coupled receptor, GPR1 (AAV58355). GPR1 is predicted to have seven transmembrane domains, which is characteristic of G protein-coupled receptors, and also contains a very large third cytoplasmic loop and a large cytoplasmic tail. The third cytoplasmic loop contains two short basic sequences; one sequence (KKRAQIG) is present at the N-terminal end of the loop and the other sequence (KKRRRAQIQ) is at the C-terminal end of the loop. The third cytoplasmic loop also contains an asparagine-rich region. GPR1 is believed to provide the upstream signal that activates GPA2, and plays an essential role in inducing the switch from non-filamentous to filamentous growth in yeast. Modulators of GPR1 function, especially inhibitors, are useful for prevention or treatment of a disease or disorder involving a fungal infection. The inhibitor may be an anti-GPR1 antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an inserted heterologous sequence inactivates the biological activity of GPR1. Compounds which modulate GPR1 gene expression and/or gene product activity may be used to inhibit the conversion of a fungus from a non-filamentous form to a filamentous form. The present sequence represents a GPR mutant, d610-617, used in an exemplification of the invention. This sequence is lacking the second basic motif (KKRAQIQ) of the third cytoplasmic loop, relative to the native GPR1. Note: This sequence is not shown in the specification, but is derived from the yeast GPR1 amino acid sequence given in figure 1A.							
XX	SQ	Sequence 953 AA;							
Query Match		85.7%; Score 30; DP 21; Length 953;							
Best Local Similarity		66.7%; Prod. No. 4.7e+02;							
Matches	4:	Conservative	1:	Mismatches	1:	Indels	0:	Gaps	0:

QY	1	FKPXWR 6							
Db	156	fkpnwk 161							
RESULT 11									
AAV58355	ID	AAV58355 standard; Protein: 961 AA.							
XX	AC	AAV58355;							
XX	DT	27-MAR-2000 (first entry)							
XX	DE	Yeast G protein-coupled receptor GPR1.							
XX	KW	Yeast: G protein-coupled receptor; GPR1; filamentous growth; pseudohyphal form; signalling pathway; antifungal.							
XX	KW	Saccharomyces cerevisiae.							
XX	Key	Location/Qualifiers							
FT	Region	1..55	/note= "Extracellular N-terminus"						
FT	Domain	56..80	/note= "Transmembrane domain 1"						
FT	Region	81..90	/note= "Intracellular loop 1"						
FT	Domain	91..115	/note= "Transmembrane domain 2"						
FT	Region	116..132	/note= "Extracellular loop 1"						
FT	Domain	133..156	/note= "Transmembrane domain 3"						
FT	Region	157..180	/note= "Intracellular loop 2"						
FT	Domain	181..200	/note= "Transmembrane domain 4"						
FT	Region	201..246	/note= "Extracellular loop 2"						
FT	Domain	247..276	/note= "Transmembrane domain 5"						
FT	Misc-difference	250	/note= "Encoded by AGC"						
FT	Region	277..620	/note= "Intracellular loop 3"						
FT	Region	277..284	/note= "Basic motif 1 (KKRAQIG)"						
FT	Region	490..586	/note= "Asparagine-rich region"						
FT	Region	610..617	/note= "Basic motif 2 (KKRAQIQ)"						
FT	Domain	621..645	/note= "Transmembrane domain 6"						
FT	Region	646..656	/note= "Extracellular loop 3"						
FT	Domain	657..678	/note= "Transmembrane domain 7"						
FT	Region	679..961	/note= "Intracellular C-terminus"						
FT	Misc-difference	250	/note= "Encoded by AGC"						
XX	WO9963094-A2.								
XX	XX	03-DEC-1999.							
XX	XX	28-MAY-1999; 99WO-US11838.							
XX	XX	01-JUN-1998; 98US-0088311.							
XX	PA	(MOUN) MOUNT SINAI SCHOOL MEDICINE.							
XX	PI	Hirsch JP, Xue Y;							

XX WPI; 2000-086980/07.
 XX N-PSDB; AA255699.
 XX Novel protein, gene, antibody and recombinant cell useful for
 PT identifying modulators and inhibitors useful as anti-fungals -
 XX
 PS Claim 4; Fig 1A; 63pp; English.
 XX This sequence represents a novel yeast G protein-coupled receptor,
 CC GPR1. GPR1 is predicted to have seven transmembrane domains,
 CC which is characteristic of G protein-coupled receptors, and also
 CC contains a very large third cytoplasmic loop and a large cytoplasmic
 CC tail. The third cytoplasmic loop contains two short basic sequences; one
 CC sequence (KKIKAQIG) is present at the N-terminal end of the loop and the
 CC other sequence (KKRRRAQIQ) is at the C-terminal end of the loop. The
 CC third cytoplasmic loop also contains an asparagine-rich region. GPR1 is
 CC believed to provide the upstream signal that activates GPA2, and plays
 CC an essential role in inducing the switch from non-filamentous to
 CC filamentous growth in yeast. Modulators of GPR1 function, especially
 CC inhibitors, are useful for prevention or treatment of a disease or
 CC disorder involving a fungal infection. The inhibitor may be an anti-GPR1
 CC antibody, GPR1 antisense nucleic acid or a GPR1 gene altered so that an
 CC inserted heterologous sequence inactivates the biological activity of
 CC GPR1. Compounds which modulate GPR1 gene expression and/or gene product
 CC activity may be used to inhibit the conversion of a fungus from a
 CC non-filamentous form to a filamentous form.
 XX Sequence 961 AA;
 SQ

Query Match 85.7%; Score 30; DB 21; Length 961;
 Best Local Similarity 66.7%; Pred. No. 4.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 OY 1 FKPNWR 6
 DB 156 fknw 161
 III I;
 RESULT 12
 AAR41001
 ID AAR41001 standard; Protein; 1093 AA.
 AC AAR41001;
 XX 25-FEB-1994 (first entry)
 DT Human myotonic dystrophy gene protein.
 DE Abnormality; muscular dystrophy; CHR 19; chromosome 19;
 KW protein kinase; polymerase chain reaction; brain.
 XX Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Region 1..1093
 FT /note= "encoded by predicted reading frame a,
 FT x's in the sequence indicate stop codons
 FT in the reading frame"
 XX
 PN W09317104-A.
 XX 02-SEP-1993.
 PD
 XX 19-FEB-1993; 93WO-US01545.
 PF
 XX 20-FEB-1992; 92US-0839255.
 PR
 XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
 PA Brook ID, Housman DE;
 XX

DR WPI; 1993-288410/36.
 XX DNA sequence of myotonic dystrophy gene - used to produce probes
 PT and identify CHR 19 abnormality and protein kinase responsible
 XX
 PS Disclosure; Fig 6; 64pp; English.
 XX The sequence is that encoded by predicted reading frame a of
 CC the human myotonic dystrophy (DM) gene. It may be used in the
 CC identification of individuals affected by DM.
 XX Sequence 1093 AA;
 SQ
 Query Match 85.7%; Score 30; DB 14; Length 1093;
 Best Local Similarity 66.7%; Pred. No. 5.5e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 OY 1 FKPNWR 6
 DB 392 fknw 387
 III I;
 RESULT 13
 AAY51233
 ID AAY51233 standard; Protein; 2204 AA.
 AC AAY51233;
 XX 07-APR-2000 (first entry)
 DT Newcastle disease virus LaSota genome encoded protein 6.
 DE Avian-paramyxovirus; infection; lentogenic; F protein; vaccing;
 KW respiratory disease; gastrointestinal disease; poultry pathogen;
 XX local immunity.
 XX Newcastle disease virus.
 OS
 XX W05966045-A1.
 PN
 XX 23-DEC-1999.
 PD
 XX 17-JUN-1999; 99WO-NL00377.
 PF
 XX 19-JUN-1998; 98EP-0202054.
 PR
 XX (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
 PA
 XX Partners BPH, De Leeuw OS, Koch G, Gielkens ALJ;
 PT
 XX WPI; 2000-106102/09.
 DR
 XX New avian paramyxovirus cDNA, useful for production of vaccine against
 PT Newcastle disease virus
 PT
 XX Disclosure; Fig 3; 115pp; English.
 PS
 XX This invention describes a novel avian-paramyxovirus cDNA (I) which
 CC comprises a nucleic acid sequence corresponding to the 5' terminal
 CC end of the genome of avian-paramyxovirus allowing the generation of
 CC an infectious copy of avian-paramyxovirus. The cell line is useful for
 CC the production of infectious lentogenic NDV (Newcastle Disease virus)
 CC without the addition of exogenous proteolytic activity. Also it is
 CC possible to generate a stable transfected cell line that expresses the
 CC wild-type F protein in the virus envelope therefore providing infectious
 CC particles, useful in the form of a vaccine, especially against
 CC respiratory and/or gastrointestinal diseases. NDV can be easily cultured
 CC to very high titers in embryonated eggs. Mass culture of embryonated
 CC eggs is relatively cheap. NDV vaccines are relatively stable and can be
 CC simply administered by mass application methods e.g. drinking water or
 CC by spraying or by aerosol formation. The natural route of infection is
 CC by the respiratory and/or gastrointestinal tract which are also the major

CC routes of infection of many other poultry pathogens, NDV can induce local
 CC immunity despite the presence of circulating maternal antibody. This
 CC sequence represents a protein encoded by the NDV strain IaSota genome
 CC which is described in the method of the invention.

XX SQ Sequence 2204 AA;
 Query Match 85.7%; Score 30; DB 21; Length 2204;
 Host Local Similarity 66.7%; Pred. No. 1.1e+03;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 I I I I
 DB 1844 fplwr 1849

RESULT 14
 AAB53989
 ID AAB53989 standard; Protein: 90 AA.

XX AC AAB53989;

XX DT 09-MAR-2001 (first entry)

XX DE Human colon cancer antigen protein sequence SEQ ID NO:1529.

XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cytostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antineoplastic; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder.

XX OS Homo sapiens.

XX PN W020005351-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000MO-US05883.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX WPI: 2000-587534/55.

XX DR N-PSDB; AAC98746.

XX PT Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer.

XX PS Claim 11; page 2083-2084; 2104pp; English.

XX CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antineoplastic, antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment, and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present

CC Invention.
 XX Sequence 90 AA;
 SQ

Query Match 82.9%; Score 29; DB 21; Length 90;
 Best Local Similarity 66.7%; Pred. No. 67;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKWR 6
 I I I I
 DB 74 fkpck 79

RESULT 15

AAB93531
 ID AAB93531 standard; Protein: 217 AA.

XX AC AAB93531;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:12886.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN E91074617-A2.

XX PD 07-FEB-2001.

XX PF 26-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI: 2001-318749/34.

XX PS Claim 8; SEQ ID 12886; 2537pp + CD ROM; English.

XX CC The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification, where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAB03166 to AAB13628 and
 CC AAB13633 to AAB18742 represent human cDNA sequences; AAB92446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 of the present invention.

XX
 SQ Sequence 217 AA;

Query Match 82.9%; Score 29; DB 22; Length 217;
 Best Local Similarity 80.0%; Pred. No. 1.7e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKW 5
 III:1
 Db 164 fkpaw 168

Search completed: February 27, 2002, 11:41:16
 Job time: 455 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:45:19 ; Search time 145.23 Seconds
(without alignments)
3.147 Million cell updates/sec

Title: US-09-446-109a-9

Perfect score: 35

Sequence: 1 FRPXR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DH seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR_68:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	380	2 C64364	formate hydrogenly
2	33	94.3	412	2 E65146	hypothetical 47.4
3	33	94.3	498	2 H71279	probable femA prot
4	33	94.3	557	2 B86020	hypothetical prote
5	33	94.3	623	2 T18250	peptide transport
6	31	88.6	1035	2 S78199	probable maturase
7	30	85.7	239	2 H72667	hypothetical prote
8	30	85.7	462	2 E65131	hypothetical 47.5
9	30	85.7	462	2 B86002	probable amino aci
10	30	85.7	606	2 H69805	conserved hypothet
11	30	85.7	684	2 T31640	hypothetical prote
12	30	85.7	961	2 S67568	probable membrane
13	30	85.7	962	2 JC3608	G protein-coupled
14	30	85.7	2204	1 RRMZNV	genome polypeptid
15	29	82.9	119	2 T36433	hypothetical prote
16	29	82.9	217	2 H64956	yedG protein - Esc
17	29	82.9	222	2 F85810	hypothetical prote
18	29	82.9	301	2 T24993	hypothetical prote
19	29	82.9	395	2 H69147	LPS biosynthesis R
20	29	82.9	412	2 H75484	hypothetical prote
21	29	82.9	435	2 C86340	protein F2D10.28 I
22	29	82.9	458	2 H82504	conserved hypothet
23	29	82.9	462	2 T15052	amino acid permeas
24	29	82.9	467	2 A59268	modulation competi
25	29	82.9	473	2 G72753	hypothetical prote
26	29	82.9	477	2 S77373	hypothetical prote
27	29	82.9	488	2 B82967	probable glucose-6
28	29	82.9	522	2 T05142	glutamate--cystein
29	29	82.9	554	1 S34607	carboxylesterase (

ALIGNMENTS

RESULT 1

C64364

Formate hydrogenlyase, subunit 5 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Data: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: C64364

R:Bult, G.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999

A:Accession: C64364

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-380 <BL>

C:Cross-references: GB:U67501; GB:L77117; NID:g2826289; PIDN:AAB98504.1; PID:g1591218

C:Genes: 13:

A:Map position: REV458767-457625

C:Superfamily: NADH dehydrogenase (ubiquinone) 49K protein

Query Match 94.3%; Score 33; DB 2; Length 380;
Best Local Similarity 83.3%; Pred. No. 27;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXR 6

DB 241 FRPVR 246

RESULT 2

E65145

hypothetical 47.4 kD protein in rhsb-pit intergenic region - Escherichia coli (strain

N:Alternate names: hypothetical protein f409

C:Species: Escherichia coli

C:Data: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999

C:Accession: E65146; S47710

K:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.; Rose, D.J.; Mau, B.; Shao, Y.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: E65146

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-412 <BLAT>

A:Cross-references: GB:AE000425; GB:U00096; NID:g2367232; PIDN:AAC76515.1; PID:g23672

A:Experimental source: strain K-12, substrain MG1655

K:Plumett, G.

Submitted to the EMBL Data Library, March 1994

Query Match 88.6%; Score 31; DB 2; Length 1035;
 Best Local Similarity 66.7%; Pred. No. 1.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 FKPKWR 6
 III I
 Db 935 FKPSWK 940

RESULT 7
 Hypothetical: protein APE0766 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: H72667
 R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takai
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
 DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
 A:Reference number: A72450; MUID:99310339
 A:Accession: H72667
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-239 <KAW>
 A:Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BA79744.1; PID:95104429
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0766
 C:Superfamily: Methanobacterium thermoautotrophicum conserved hypothetical protein MTU10

Query Match 85.7%; Score 30; DB 2; Length 239;
 Best Local Similarity 66.7%; Pred. No. 70;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 FKPKWR 6
 III I
 Db 105 FKPSWQ 110

RESULT 8
 Hypothetical 47.5 kD protein in cysG-trpS intergenic region - Escherichia coli (strain K
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 17-Mar-2000
 C:Accession: E65131
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617
 A:Accession: E65131
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-462 <BLAT>
 A:Cross-references: GB:AE000413; GB:U00096; NID:92367215; PIDN:AAC76395.1; PID:92367216;
 A:Experimental source: strain K-12, substrain MG1655
 C:Genetics:
 A:Gene: ynfM
 C:Superfamily: arginine permease

Query Match 85.7%; Score 30; DB 2; Length 462;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 FKPKWR 6
 III I
 Db 398 YKPLWR 403

RESULT 9
 Hypothetical protein Y57A10A.r - Caenorhabditis elegans

B86002
 probable amino acid/amine transport protein ynfM [imported] - Escherichia coli (strain
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
 C:Accession: B86002
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 sler, G.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Polamoukis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: B86002
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-462 <STO>
 A:Cross-references: GB:AE005174; NID:912517998; PIDN:AA058478.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ynfM

Query Match 85.7%; Score 30; DB 2; Length 462;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 FKPKWR 5
 III I
 Db 398 YKPLWR 403

RESULT 10
 B69803
 Conserved hypothetical protein ynfX - Bacillus subtilis
 C:Species: Bacillus subtilis
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
 C:Accession: B69803
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 393, 249-256, 1997
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
 A:Reference number: A69580; MUID:98044033
 A:Accession: B69803
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-606 <KUN>
 A:Cross-references: GB:299108; GB:AL009126; NID:92633055; PIDN:CAB12672.1; PID:e11828
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: ynfX

Query Match 85.7%; Score 30; DB 2; Length 606;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 FKPKWR 5
 III I
 Db 570 YKPEWR 575

RESULT 11
 T31643
 hypothetical protein Y57A10A.r - Caenorhabditis elegans

C:Species: *Caenorhabditis elegans*
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T31640
 R:Smyle, R.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: 221048
 A:Accession: T31640
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-684 <WIL>
 A:Cross-references: EMBL:AL117195; NID:e1549729; PIDN:CAB55023.1; CESP:Y57A10A.r
 A:Experimental source: clone Y57A10A
 C:Genetics:
 A:Gene: CESP:Y57A10A.r
 A:Introns: 108/3; 132/3; 149/1; 180/3; 323/2; 430/1; 530/1; 584/2; 647/1

Query Match 85.7%; Score 30; DB 2; Length 684;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 III I:
 DB 672 FKPKWK 677

RESULT 12
 S67568
 Probable membrane protein YDL035c - yeast (*Saccharomyces cerevisiae*)
 N:Alternate names: hypothetical protein D2749
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 29-Oct-1999
 C:Accession: S67568
 R:Paulin, L.; Saron, A.M.; Laamanen, V.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67560
 A:Accession: S67568
 A:Molecule type: DNA
 A:Residues: 1-961 <PAU>
 A:Cross-references: EMBL:Z74083; NID:g1431014; PID:e252988; PID:g1431015; GSPDB:GNO0004.
 A:Experimental source: strain S28C
 C:Genetics:
 A:Gene: MIPS:YDL035c
 A:Map position: 4L
 C:Keywords: transmembrane protein
 F:57-73/Domain: transmembrane #status predicted <TM1>
 F:92-108/Domain: transmembrane #status predicted <TM2>
 F:140-156/Domain: transmembrane #status predicted <TM3>
 F:181-197/Domain: transmembrane #status predicted <TM4>
 F:256-272/Domain: transmembrane #status predicted <TM5>
 F:623-639/Domain: transmembrane #status predicted <TM6>
 F:662-678/Domain: transmembrane #status predicted <TM7>

Query Match 85.7%; Score 30; DB 2; Length 961;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 III I:
 DB 156 FKPNWK 161

RESULT 13
 JC5808
 G protein-coupled receptor 1 - yeast (*Saccharomyces cerevisiae*)
 C:Species: *Saccharomyces cerevisiae*
 C:Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 19-May-2000
 C:Accession: JC5808
 R:Yun, C.W.; Yamaki, H.; Nakayama, R.; Yamamoto, K.; Kumagai, H.
 Biochem. Biophys. Res. Commun. 240, 287-292, 1997
 A:Title: G-protein coupled receptor from yeast *Saccharomyces cerevisiae*.
 A:Reference number: JC5808; NUID:98049822

A:Accession: JC5808
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-962 <YUN>
 C:Comment: This protein monitors the extracellular signal such as nutrition and trans

C:Genetics:
 A:Gene: gpr1
 F:56-74/Domain: transmembrane #status predicted <TM1>
 F:91-105/Domain: transmembrane #status predicted <TM2>
 F:139-157/Domain: transmembrane #status predicted <TM3>
 F:180-198/Domain: transmembrane #status predicted <TM4>
 F:255-273/Domain: transmembrane #status predicted <TM5>
 F:622-640/Domain: transmembrane #status predicted <TM6>
 F:661-679/Domain: transmembrane #status predicted <TM7>

Query Match 85.7%; Score 30; DB 2; Length 962;
 Best Local Similarity 66.7%; Pred. No. 2.6e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 III I:
 DB 156 FKPNWK 161

RESULT 14
 R3NZNV
 Genome polypeptide - Newcastle disease virus (strain Beaudette C)
 N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)
 C:Species: Newcastle disease virus
 C:Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 11-Jun-1999
 C:Accession: A26747
 R:Yusoff, K.; Millar, N.S.; Chambers, P.; Emmerson, P.T.
 Nucleic Acids Res. 15, 3961-3976, 1987
 A:Title: Nucleotide sequence analysis of the L gene of Newcastle disease virus: homolog
 A:Reference number: A93665; NUID:87230982
 A:Accession: A26747
 A:Molecule type: mRNA
 A:Residues: 1-2204 <YUS>
 A:Cross-references: GB:X05399; NID:g60937; PIDN:CAA28985.1; PID:g60939
 C:Genetics:
 A:Gene: L
 C:Superfamily: parainfluenza virus RNA-directed RNA polymerase
 C:Keywords: Atp; nucleotidyltransferase

Query Match 85.7%; Score 30; DB 1; Length 2204;
 Best Local Similarity 66.7%; Pred. No. 5.8e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 I: I I:
 DB 1344 FRPLWR 1849

RESULT 15
 T36433
 Hypothetical protein SCF43A.10c - *Streptomyces coelicolor*
 C:Species: *Streptomyces coelicolor*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T36433
 R:Seeyer, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, July 1999
 A:Reference number: 221598
 A:Accession: T36433
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-119 <SEE>
 A:Cross-references: EMBL:AL096837; PIDN:CAB48897.1; GSPDB:GNO0070; SCOEDB:SCF43A.10c
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SCF43A.10c

Query Match 82.9%; Score 29; DB 2; Length 119;
Best Local Similarity 66.7%; Pred. No. 57;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
I: I I I
DB 36 FEPCWR 41

Search completed: February 27, 2002, 11:45:19
Job time: 698 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:42:45 ; Search time 78.39 Seconds
(without alignments)
2.806 Million cell updates/sec

Title: US-09-446-109A-9
Perfect score: 35
Sequence: 1 FKXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSPROT_39.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	535	1 YHIL_ECOLI	P37629 escherichia
2	33	94.3	623	1 PTR2_CANAL	P46030 candida alb
3	31	88.6	807	1 YM91_SCHPO	P05511 schizosacch
4	30	85.7	445	1 YHFW_ECOLI	P45539 escherichia
5	30	85.7	2204	1 RRPL_NDVB	P11205 newcastled
6	29	82.9	217	1 YEDK_ECOLI	P76318 escherichia
7	29	82.9	452	1 SPS2_MOUSE	P97364 mus musculu
8	29	82.9	477	1 YZ6A_SYNY3	P73436 synchoocyst
9	29	82.9	514	1 GSHI_BRAJU	O23736 brassica ju
10	29	82.9	522	1 GSHI_ARATH	P46309 arabidopsis
11	29	82.9	554	1 ESTM_MOUSE	Q63880 mus musculu
12	28	80.0	151	1 HSPD_BRAJA	O69241 bradyrhizob
13	28	80.0	151	1 HSPH_BRAJA	O86110 bradyrhizob
14	28	80.0	162	1 YLF6_CABEL	O03598 caenorhabdi
15	28	80.0	167	1 GCG6_HUMAN	Q9Y3C8 homo sapien
16	28	80.0	204	1 DSBA_LEGPN	P50024 legionella
17	28	80.0	233	1 VENV_BEV	P27904 berne virus
18	28	80.0	307	1 CYPE_DROME	Q9V3G3 drosophila
19	28	80.0	337	1 YHLD_ECOLI	P37642 escherichia
20	28	80.0	340	1 C5AK_CORG0	P79175 gorilla gor
21	28	80.0	340	1 C5AK_PANTH	P79240 pan troglod
22	28	80.0	347	1 C5AR_MOUSE	P30993 mus musculu
23	28	80.0	350	1 C5AR_HUMAN	P21730 homo sapien
24	28	80.0	352	1 C5AR_RAT	P97520 rattus norv
25	28	80.0	388	1 DXR_BACSU	O31753 bacillus su
26	28	80.0	405	1 HS47_CHICK	P13731 gallus gall
27	28	80.0	417	1 HS47_HUMAN	P29043 homo sapien
28	28	80.0	417	1 HS47_MOUSE	P19324 mus musculu
29	28	80.0	417	1 HS47_RAT	P29457 rattus norv
30	28	80.0	418	1 CDP2_HUMAN	P50454 homo sapien
31	28	80.0	469	1 RDXA_RHOSH	Q01854 rhodobacter
32	28	80.0	482	1 C3AR_HUMAN	Q16581 homo sapien
33	28	80.0	482	1 PUR8_YEAST	Q05911 saccharomyc

34 28 80.0 490 1 C883_ARATH
35 28 80.0 501 1 YDF7_SCHPO
36 28 80.0 508 1 GLPK_MYCLE
37 28 80.0 639 1 V70K_PLRV1
38 28 80.0 843 1 POL_MLVAK
39 28 80.0 901 1 PODK_TREPA
40 28 80.0 909 1 YAGI_RHISN
41 28 80.0 967 1 Y5G0_CLOAB
42 28 80.0 982 1 HBL1_CAEEL
43 28 80.0 1133 1 ATX9_TETTH
44 28 80.0 1196 1 POL_MLVAV
45 28 80.0 1275 1 RFBC_MYXXA

O23051 arabidopsis
O10479 schizosacch
Q9CB81 mycobacteri
P17519 potato leaf
P03357 akr murine
O83728 treponema p
P55465 rhizobium s
P33747 clostridium
Q9XYD3 caenorhabdi
Q95050 tetrahymena
P03356 akr murine
Q50864 myxococcus

ALIGNMENTS

RESULT 1
YHIL_ECOLI STANDARD; PRT: 535 AA.
AC P37629; P37628;
DT 01-OCT-1994 (Rel. 30, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 61.6 KDA PROTEIN IN RHB-PIT INTERGENIC REGION.
GN YHIL OR B3489/B3490.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN-R12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT fusion from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
EN 2;
RP REVISIONS.
RC STRAIN-R12 / MG1655;
RX MEDLINE=97425617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Peria N.T., Burland V.,
RA Davis M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregory J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shan Y.;
RL "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RC CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL "Published observations (JUN-1999).
CC CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC INTRODUCED IN POSITION 399 TO PRODUCE THIS ORF.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U00039; AAB18465.1; ALT_FRAME.
CC EMBL: U00039; AAB18466.1; ALT_FRAME.
CC EMBL: AE000425; AAC76514.1; ALT_FRAME.
CC EMBL: AE000425; AAC76515.1; ALT_FRAME.
CC Ecogene: EG12227; ynil.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 535 AA: 61576 MW: 0046CE53998989DF1 CRC64;

Query: Match 94.3% Score 33; DB 1; Length 535;


```

FT ACT_SITE 63 63 POTENTIAL..
FT SE_CYS 63 63
FT SITE 66 66
FT IMPORTANT FOR CATALYTIC ACTIVITY (BY
FT SIMILARITY).
FT NP_BIND 322 328
FT DOMAIN 2 9
FT DOMAIN 433 440
FT POLY-ALA.
FT POLY-ALA.
SQ SEQUENCE 452 AA: 47786 MW: 9DA6F7250CFE80E4 CRC64;

Query Match 82.9%; Score 29: DB 1: Length 452;
Best Local Similarity 66.7%; Pred.No. 1.1e+02;
Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKPxWR 6
   | | | |
Db 48 FSPSWR 53

RESULT: 8
YZ64_SYNY3
ID YZ64_SYNY3 STANDARD: PRT: 477 AA.
AC P73436;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 54.0 KDA PROTEIN SLL1464.
GN SLL1464.
OS Synechocystis sp. (strain PCC 6803).
OX Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OC NCBI_TaxID=1148;
KN | | |
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kameko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugisawa M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
EL DNA Res. 3:109-136(1996).
CC
CC -1- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90906; BAA17476.1; ..
DR InterPro: IPR003846; UPF0061.
DR Pfam: PF02696; UPF0061; 1.
DR Hypothetical protein: Complete proteome.
SQ SEQUENCE 477 AA: 54041 MW: 81F6899D1A6D613C CRC64;

Query Match 82.9%; Score 29: DB 1: Length 477;
Best Local Similarity 66.7%; Pred.No. 1.2e+02;
Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKPxWR 6
   | | | |
Db 348 FSPSWR 393

RESULT 9
GSH1_BRAJU
ID GSH1_BRAJU STANDARD: PRT: 514 AA.
AC Q23736; Q43389;

```


15-DEC-1998 (Rel. 37, Last sequence update)
 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GLUTAMATE--CYSTEINE LIGASE, CHLOROPLAST PRECURSOR (EC 6.3.2.2) (GAMMA-
 DE GLUTAMYL-CYSTEINE SYNTHETASE) (GAMMA-ECS) (GCS).
 GS1.
 OS Brassica juncea (leaf mustard) (Indian mustard).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3707;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Root;
 RX MEDLINE=98281577; PubMed=9620267;
 RA Schaefer H.J., Haag-Kerwer A., Rausch T.;
 RT "cDNA cloning and expression analysis of genes encoding GSH synthesis
 RT in roots of the heavy-metal accumulator Brassica juncea L.: evidence
 RT for Cd-induction of a putative mitochondrial gamma-glutamylcysteine
 RT synthetase isoform.";
 RT plant Mol. Biol. 37:87-97(1998).
 RN [2]
 RP SEQUENCE OF 126-342 FROM N.A.
 RC STRAIN=CV. VITTASSO; TISSUE=Leaf;
 RX MEDLINE=97227950; PubMed=9119067;
 RA Schaefer H.J., Greiner S., Rausch T., Haag-Kerwer A.;
 RT "In seedlings of the heavy metal accumulator Brassica juncea, Cu2+
 RT differentially affects transcript amounts for gamma-glutamylcysteine
 RT synthetase (gamma-ECS) and metallothionein (MT2).";
 RL FEBS Lett. 404:216-220(1997).
 CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + L-CYSTEINE -> ADP +
 CC ORTHOPHOSPHATE + GAMMA-L-GLUTAMYL-L-CYSTEINE.
 CC -!- PATHWAY: FIRST AND RATE-LIMITING STEP IN DE NOVO GLUTATHIONE
 CC BIOSYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: CHLOROPLAST (BY SIMILARITY).
 CC -!- MISCELLANEOUS: IN PLANTS, THERE ARE TWO FORMS OF THIS ENZYME, ONE
 CC CYTOSOLIC AND THE OTHER CHLOROPLASTIC.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y10848; CAA71801.1;
 DR EMBL: X95563; CAA64808.1;
 KW Glutathione biosynthesis; Ligase; Chloroplast; Transit peptide.
 FT TRANSIT 1 55 CHLOROPLAST (POTENTIAL).
 FT CHAIN 56 514 GLUTAMATE--CYSTEINE LIGASE.
 FT CONFLICT 139 139 K -> R (IN REF. 2).
 FT CONFLICT 215 215 T -> I (IN REF. 2).
 FT CONFLICT 257 257 M -> T (IN REF. 2).
 SQ SEQUENCE 514 AA: 57903 MW: 07C71CB13E785FA8 CRC64;

 Query Match 82.9%; Score 29; DB 1; Length 514;
 Best Local Similarity 66.7%; Pred: No. 1.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;

 QY 1 FKPKWR 6
 I: I I I
 DB 204 FQPKWR 209

 RESULT 10
 GSH1_ARATH STANDARD: PRT: 522 AA.
 AC P46309; O82759; P92951;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE DE GLUTAMATE--CYSTEINE LIGASE, CHLOROPLAST PRECURSOR (EC 6.3.2.2) (GAMMA-
 DE GLUTAMYL-CYSTEINE SYNTHETASE) (GAMMA-ECS) (GCS).
 GS1 OR AT4G23100 OR F7H19.290.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA; TISSUE=Leaf;
 RX MEDLINE=95023984; PubMed=7937837;
 RA May M.J., Leaver C.J.;
 RT "Arabidopsis thaliana gamma-glutamylcysteine synthetase is
 RT structurally unrelated to mammalian, yeast, and Escherichia coli
 RT homologs.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:10059-10063(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. LANDSBERG ERECTA;
 RA Ullmann P., Gondet L., Bach T.J.;
 RT "Isolation of an Arabidopsis thaliana cDNA encoding a putative
 RT gamma-glutamylcysteine synthetase by complementation of a GSH1
 RT deficient yeast mutant-glutamylcysteine synthetase.";
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Cobbett C.S., May M.J., Howden R., Rolfs B.;
 RT "The glutathione-deficient, cadmium-sensitive mutant, cad2-1, of
 RT Arabidopsis thaliana is deficient in gamma-glutamylcysteine
 RT synthetase.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Anisorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaler B., Mache R., Mueller M.,
 RA Krels M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.A., McCullagh B., Bligham L., Robben J.,
 RA Van der Schueren J., Grymoprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Waltjens I., Voet M., Bastiaens I., Hilbert H., Braun M.,
 RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Molijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Bernauer S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA de Keyser A., Buysschaert C., Glien J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Douquet J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
 RA Pettett A., Rajandream M.-A., Lyne M., Benes V., Rehmann S.,
 RA Horkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Drees S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Arquier A., Vitale D., Liguori R., Piravandi E.,
 RA Mussonnet O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schinagl S., Hiller R., Schmidt W., Lechamy A., Aubourg S.,
 RA Chetoui F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Reijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 RA Frishman D., Haase D., Lemcke K., Meves H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threiden J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Lefeville P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kizner J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,

```

RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tull S.,
RA Granat S., Shohdy N., Hasegawa A., Iiamsed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.,
RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana."
RI
CC Nature 402:769-777(1999).
CC
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + L-CYSTEINE -> ADP +
CC ORTHOPHOSPHATE + GAMMA-L-GLUTAMYL-L-CYSTEINE.
CC
CC -1- ENZYME REGULATION: FEEDBACK INHIBITION BY GLUTATHIONE.
CC
CC -1- PATHWAY: FIRST AND RATE-LIMITING STEP IN DE NOVO GLUTATHIONE
CC BIOSYNTHESIS.
CC
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST (PROBABLE).
CC
CC -1- TISSUE SPECIFICITY: ABUNDANT IN LEAVES AND ROOTS.
CC
CC -1- MISCELLANEOUS: IN PLANTS, THERE ARE TWO FORMS OF THIS ENZYME, ONE
CC CYTOSOLIC AND THE OTHER CHLOROPLASTIC.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z29490; CAA82626.1;
CC EMBL: Y09944; CAA71075.1;
CC EMBL: AF068299; AAD14344.1;
CC EMBL: AL031018; CAA19826.1;
CC EMBL: AL161558; CAA79265.1;
CC EMBL: 17259; Arach.2621.17259.
CC EMBL: 33329; Arach.2621.33329.
CC EMBL: 36232; Arach.2621.36232.
CC
CC Glutathione biosynthesis; Ligase; Chloroplast; Transit peptide.
CC TRANSIT ? ? ? CHLOROPLAST (POTENTIAL).
CC CHAIN 1 522 GLUTAMATE--CYSTEINE LIGASE.
CC VARIANT 22 22 A -> T (IN STRAIN LANDSBERG BRECTA).
CC VARIANT 48 48 Y -> N (IN STRAIN LANDSBERG BRECTA).
CC CONFLICT 491 522 VRTGVTPAEKLLLENGECSQSDPVFEELLY -> GQNRN
CC YACEALGDGVQWRMGTRRSRVRRAAVLRKWDVNRKCL
CC (IN REF. 1).
CC
CC SEQUENCE 522 AA; 58562 MW; CCBF13C6F44E0EF7 CRC64;
CC
CC
CC Query Match 82.9%; Score 29; DB 1; Length 522;
CC Best Local Similarity 66.7%; Pred. No. 1.3e+02;
CC Matches 4; Conservative 1; Mismatches 1; Indels 0; Caps 0;
CC
CC QY 1 FKPKXR 6
CC 1:1 11
CC DB 212 FQPKWR 217
CC
CC RESULT 11
CC ESTM_MOUSE STANDARD; PRT: 554 AA.
CC AC Q63880;
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, last sequence update)
CC DT 01-NOV-1997 (Rel. 35, last annotation update)
CC DE LIVER CARBOXYLESTERASE PRECURSOR (EC 3.1.1.1) (FS-MALE) (ESTERASE-31).
CC GN ES31.
CC OS Mus musculus (Mouse).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC OX NCBI_TaxID=10090;
CC RN 11
CC RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
CC RC STRAIN-DALBY/CJ X DBA/2J; TISSUE=Liver;

```

```

RX MEDLINE-93326638; PubMed-7916639;
RA Alda K., Moore R., Negishi M.;
RT "Cloning and nucleotide sequence of a novel, male-predominant
RT carboxylesterase in mouse liver.";
RI Blochim. Biophys. Acta 1174:72-74(1993).
CC
CC -1- FUNCTION: INVOLVED IN THE DETOXIFICATION OF XENOBIOTICS AND IN
CC THE ACTIVATION OF ESTER AND AMIDE PRODRUGS.
CC -1- CATALYTIC ACTIVITY: A CARBOXYLIC ESTER + H(2)O -> AN ALCOHOL
CC + A CARBOXYLIC ANION.
CC -1- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE, LUMEN OF ENDOPLASMIC
CC RETICULUM.
CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: S64130; AAB27606.1;
CC EMBL: P21836; IMAH.
CC MGD: MGI:102773; ES31.
CC InterPro: IPR002018; Carboxylesterase_B.
CC InterPro: IPR000379; Est_lip_thioest_actile.
CC Pfam: PF00135; Coesterase_1.
CC PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
CC PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
CC Glycoprotein; Hydrolase; Serine esterase; Endoplasmic reticulum;
CC Signal; Multigene family.
CC SIGNAL 1 14
CC CHAIN 15 554 BY SIMILARITY.
CC ACT_SITE 215 215 LIVER CARBOXYLESTERASE.
CC ACT_SITE 443 443 BY SIMILARITY.
CC DISULFID 83 110 BY SIMILARITY.
CC DISULFID 267 278 BY SIMILARITY.
CC SITE 551 554 PREVENT SECRETION FROM ER (POTENTIAL).
CC CARBOHYD 294 294 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 300 300 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 381 381 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 554 AA; 61509 MW; 028D898B06F3CAAD CRC64;
CC
CC Query Match 82.9%; Score 29; DB 1; Length 554;
CC Best Local Similarity 80.0%; Pred. No. 1.3e+02;
CC Matches 4; Conservative 0; Mismatches 1; Indels 0; Caps 0;
CC
CC QY 1 FKPKXR 5
CC 1:1 1
CC DB 434 FKPAW 438
CC
CC RESULT 12
CC HSPD_BRAJA STANDARD; PRT: 151 AA.
CC AC O69241;
CC DT 30-MAY-2000 (Rel. 39, Created)
CC DT 30-MAY-2000 (Rel. 39, last sequence update)
CC DT 30-MAY-2000 (Rel. 39, last annotation update)
CC DE SMALL HEAT SHOCK PROTEIN HSPD.
CC GN HSPD.
CC OS Bradyrhizobium japonicum.
CC OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC OC Bradyrhizobium group; Bradyrhizobium.
CC OX NCBI_TaxID=375;
CC RN 11
CC RP SEQUENCE FROM N.A.
CC RA Warberhaus F., Weiglhofer W., Fischer H.M., Hennecke H.;
CC RT "(Identification of the Bradyrhizobium japonicum desp gene as part of
CC an operon containing small heat shock protein genes.";
CC RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC
CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)

```

FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ003064; CAA05835.1; .
 DR InterPro: IPR002068; Crystallin_HSP20.
 DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SO SEQUENCE 151 AA; 17272 MW; FB44EF94F8599BE4 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 64;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKPKXR 6
 DB 6 FSPLWR 11

RESULT 13
 HSPH_BRAJA STANDARD; PRT; 151 AA.
 AC 086110;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SMALL HEAT SHOCK PROTEIN HSPH.
 GN HSPH.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 KN 11
 RP SEQUENCE FROM N.A.
 RA Muenchbach M., Nocker A., Narberhaus F.;
 RT "Occurrence of a superfamily of small heat shock proteins in
 RT Bradyrhizobium japonicum and other Rhizobium species: a plant-like
 RT phenomenon."
 RI Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AJ010144; CAA09014.1; .
 DR InterPro: IPR002068; Crystallin_HSP20.
 DR Pfam: PF00011; HSP20; 1.
 DR PROSITE: PS01031; HSP20; 1.
 KW Heat shock; Multigene family.
 SO SEQUENCE 151 AA; 17098 MW; DFE9809310AF6A28 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 151;
 Best Local Similarity 66.7%; Pred. No. 64;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FKPKXR 6
 DB 6 FSPLWR 11

RESULT 14
 YLF6_CABEL STANDARD; PRT; 162 AA.
 AC 003598;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE HYPOTHETICAL 18.5 KDA PROTEIN C40H1.6 IN CHROMOSOME 111.
 GN C40H1.6.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 CC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 KN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin K., Favello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kersey J., Kirsten J., Laister N.,
 RA LaRuelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,
 RA Sims M., Smalton N., Smith A., Smith M., Sonnhammer E., Staden R.,
 RA Sturton J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wyldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RI Nature 368:32-38(1994).
 CC -1- SIMILARITY: STRONG, TO HUMAN CGI-126.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z19154; CAA79557.1; .
 DR PIR: S28301; S28301.
 DR Wormpep: C40H1.6; CE00114.
 KW Hypothetical protein.
 CC SEQUENCE 162 AA; 18537 MW; 8DC03CE3BCE79D55 CRC64;

Query Match 80.0%; Score 28; DB 1; Length 162;
 Best Local Similarity 80.0%; Pred. No. 68;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPKXR 5
 DB 120 FKPLW 124

RESULT 15
 GGC6_HUMAN STANDARD; PRT; 167 AA.
 AC 003598;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN CGI-126.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 KN 11
 RP SEQUENCE FROM N.A.

```

RX MEDLINE=20272150; PubMed=10810093;
RA Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
RT "Identification of novel human genes evolutionarily conserved in
RL Caenorhabditis elegans by comparative proteomics.";
RL Genome Res. 10:703-713(2000).
CC -1- SIMILARITY: STRONG, TO C.ELEGANS C40H1.6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified, and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF151804; AAD34121.1; -
KW Hypothetical protein
SQ SEQUENCE 167 AA: 19458 MW: 1675D9187DC43E14 CRC64;

```

```

Query Match      80.0%; Score 28; DB 1; Length 167;
Best Local Similarity 80.0%; Pred. No. 70;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 FRPXW 5
Db 121 FRPLW 125

```

```

Search completed: February 27, 2002, 11:42:45
Job time: 544 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM' protein - protein search, using sw model

Run on: February 27, 2002, 11:50:09 ; Search time 281.76 Seconds
(without alignments)
3.115 Million cell updates/sec

Title: US-09-446-109a-9
Perfect score: 35
Sequence: 1 FKXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : SPTREMBL17:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp Vertebrate:*
 - 14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	97.1	250	2 Q9EUM2	Q9eum2 corynebacte
2	33	94.3	380	1 Q57935	Q57935 methanococc
3	33	94.3	498	2 Q83784	Q83784 treponema p
4	33	94.3	623	3 Q9URL7	Q9url7 candida alb
5	33	94.3	917	5 Q9XYP8	Q9xyp8 drosophila
6	33	94.3	917	5 Q9VY94	Q9vy94 drosophila
7	31	88.6	221	10 Q9XJ12	Q9xj12 oryza sativ
8	30	85.7	204	11 Q9D3F2	Q9d3f2 mus musculu
9	30	85.7	239	1 Q9YE03	Q9ye03 aeropyrum p
10	30	85.7	440	11 Q9D9R1	Q9d9r1 mus musculu
11	30	85.7	606	2 Q31566	Q31566 bacillus su
12	30	85.7	610	2 Q52961	Q52961 bacillus su
13	30	85.7	684	5 Q9NA80	Q9na80 caenorhabdi
14	30	85.7	961	3 Q12361	Q12361 saccharomyc
15	30	85.7	1232	10 Q9LJ01	Q9lj01 arabidopsis
16	30	85.7	2204	12 Q90341	Q90341 newcastle d
17	30	85.7	2204	12 Q9WMH6	Q9wmh6 newcastle d
18	30	85.7	2204	12 Q9DLJ3	Q9dlj3 newcastle d
19	30	85.7	2454	3 Q9UVP2	Q9uvp2 emericella

20	30	95.7	2454	3 Q9UUV56	Q9uv56 emericella
21	29	82.9	119	2 Q9XAA5	Q9xaa5 streptomyce
22	29	82.9	138	2 Q9ZEW4	Q9zew4 klebsiella
23	29	82.9	143	5 Q96402	Q96402 plasmodium
24	29	82.9	162	2 Q9CJP0	Q9cjp0 pasteurella
25	29	82.9	164	5 Q9NJY1	Q9njyl giardia lam
26	29	82.9	224	5 Q9N2Y6	Q9n2y6 caenorhabdi
27	29	82.9	233	2 Q9F2R9	Q9f2r9 streptomyce
28	29	82.9	244	10 Q9FS78	Q9fs78 triticum ae
29	29	82.9	262	5 Q9NND2	Q9nnd2 leishmania
30	29	82.9	296	2 Q33353	Q33353 mycobacteri
31	29	82.9	301	5 Q22568	Q22568 caenorhabdi
32	29	82.9	395	1 Q26470	Q26470 methanobact
33	29	82.9	398	2 Q9A710	Q9a710 caulobacter
34	29	82.9	412	2 Q9RWE3	Q9rwe3 deinococcus
35	29	82.9	435	10 Q9LM75	Q9lm75 arabidopsis
36	29	82.9	436	10 Q82665	Q82665 brassica ju
37	29	82.9	445	3 Q9C1B2	Q9clb2 gibberella
38	29	82.9	447	3 Q94088	Q94088 fusarium sp
39	29	82.9	447	3 Q9C1B9	Q9clb9 fusarium sp
40	29	82.9	458	2 Q9KN89	Q9kn89 vibrio chol
41	29	82.9	462	10 Q40414	Q40414 nicotiana s
42	29	82.9	467	2 Q52729	Q52729 rhizobium e
43	29	82.9	473	1 Q9YG81	Q9yg81 aeropyrum p
44	29	82.9	488	2 Q9HTC7	Q9htc7 pseudomonas
45	29	82.9	557	3 Q42902	Q42902 schizosacch

ALIGNMENTS

RESULT 1
Q9EUM2
ID Q9EUM2 PRELIMINARY: PRT: 250 AA.
AC Q9EUM2
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Q9EUM2
GN Q9EUM2
CS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacterota: Firmicutes: Actinobacteria: Actinobacteridae:
OC Actinomycetales: Corynebacterineae: Corynebacteriaceae;
JC Corynebacterium.
UX XREF: TaxID=118;
RN 111
RP SOURCE FROM N.A.
RC SPTREMBL-ATCC3130;
FA Tsuchi A., Puchler A., Kalinowski J.;
XT rDNA sequence and genetic organization of the integron-carrying R-
ET plasmid pCG4 of Corynebacterium glutamicum.*;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DE Q9UUV56: AFI64956: AAG00294.1;
DR RefSeq: IP8002145; CopG_HTH_4.
DR RefSeq: PF01492; HTH_4; 1;
KW pCG4.
SQ SEQUENCE 259 AA: 28804 MW: 195C8G367B483157 CRC64;

Query Match 97.1% Score 34; DB 2; Length 250;
Best Local Similarity 83.3%; Pred. No. 35;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKXWR 6
DB 214 FKPAWR 219

RESULT 2
Q57935
ID Q57935 PRELIMINARY: PRT: 380 AA.
AC Q57935

DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PUTATIVE FORMATE HYDROGENLYASE SUBUNIT 5 (FHL SUBUNIT 5) (HYDROGENASE-
 DE 3 COMPONENT E).
 GN M00515.
 OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota: Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.D., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.
 RL Science 273:1058-1073(1996).
 CC -1- COFACTOR: NICKEL (BY SIMILARITY).
 CC -1- PATHWAY: HYDROGEN METABOLISM; FHL PATHWAY (BY SIMILARITY).
 CC -1- SUBUNIT: FHL COMPRISIS OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
 CC ELECTRON CARRIERS AND A HYDROGENASE (ISOENZYME 3). IN THIS NON-
 CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBODIOXIDE FROM
 CC FORMATE ARE RELEASED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 49 KDA SUBUNIT FAMILY.
 DR EMBL: U67501; AA898504.1;
 DR TIGR: M00515;
 DR InterPro: IPR001135; Complex1_49kd.
 DR InterPro: IPR001501; Nifese_Hases.
 DR Pfam: PF00346; complex1_49kd; 1.
 DR Pfam: PF00374; Nifese_Hases; 1.
 DR PROSITE: PS00535; COMPLEX1_49K; UNKNOWN_1.
 DR PROSITE: PS00507; N1_HGENASE_L1; UNKNOWN_1.
 KW Hypothetical protein: Oxidoreductase; NAD; Iron-sulfur; 4Fe-4S;
 KW Nickel; Complete proteome.
 SQ SEQUENCE 380 AA; 43638 MW; 3EAD16D67B3476F CRC64;

Query Match 94.3%; Score 33; DB 1; Length 380;
 Best Local Similarity 83.3%; Pred. No. 81;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6

DB 241 FKPVWR 246

RESULT 3

ID 083784

AC 083784 PRELIMINARY: PRT: 498 AA.

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE FHL PROTEIN, PUTATIVE.

GN TP0806.

OS Treponema pallidum.

OC Bacteria: Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI_TaxID=160;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NICHOIS.

RX MEDLINE=98332770; PubMed=9665876;

RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Sodergren E., Iardham J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Ullerback T.,
 RA McDonald L., Artlach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT *Complete genome sequence of Treponema pallidum, the syphilis
 agent, Spirochete.
 RL Science 281:375-388(1998).
 RX EMBL: AF001251; AAC65773.1;
 DR TIGR: TP0806;
 DR InterPro: IPR003447; FemAB.
 DR Pfam: PF02388; FemAB; 1.
 KW Complete proteome.
 SQ SEQUENCE 498 AA; 56074 MW; 81710E41264431B8 CRC64;

Query Match 94.3%; Score 33; DB 2; Length 498;
 Best Local Similarity 83.3%; Pred. No. 1e+02;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 DB 240 FKPVWR 245

RESULT 4

ID 090RL7

AC 090RL7 PRELIMINARY: PRT: 623 AA.

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE PEPTIDE TRANSPORT PROTEIN.

GN PTR2.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

OX NCBI_TaxID=5476;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1161;

RA Barrell B.G., Rajandream M.A.;

RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RN 12;

RP SEQUENCE FROM N.A.

RA Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RN 13;

RP SEQUENCE FROM N.A.

RC STRAIN=1161;

RX MEDLINE=97435544; PubMed=9290243;

RA Tait E., Simon M.C., King S., Brown A.J., Gow N.A.R., Shaw D.J.;

RT "A Candida albicans Genome Project: Cosmid Contigs, Physical Mapping,
 and Gene Isolation."

RL Fungal Genet. Biol. 21:308-314(1997).

RX EMBL: AL033503; CAA22021.1;

DR InterPro: IPR00109; PTR2.

DR Pfam: PF00854; PTR2; 1.

DR PROSITE: PS01022; PTR2_1; 1.

DR PROSITE: PS01023; PTR2_2; UNKNOWN_1.

SQ SEQUENCE 623 AA; 69943 MW; 8543A3D1F7E7363E CRC64;

Query Match

Best Local Similarity 94.3%; Score 33; DB 3; Length 623;

Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6

DB 445 FKPIWR 450

RESULT 5

Q9XYP8
ID Q9XYP8 PRELIMINARY: PRT: 917 AA.
AC Q9XYP8
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GAMMA-TUBULIN KINase PROTEIN DCRIP91.
GN L(1)DD4 OR DCRIP91 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila;
OX NCBI_TaxID=7227;
RN [1]
RP MEDLINE=99156983; PubMed=10037793;
RX Oegema K., Wiese C., Martin O., Milligan R.A., Iwamatsu A.,
RA Mitchison T.J., Zheng Y.;
RT "Characterization of two related Drosophila gamma-tubulin complexes
that differ in their ability to nucleate microtubules.";
RL J. Cell Biol. 144:721-733(1999).
DR EMBL: AF118380; AAD27817.1;
DR FlyBase: FBgn0001612; l(1)dd4.
DR InterPro: IPR000634; dehydrtse_ser_thr.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 917 AA; 103819 MW; EBF2F41E35049D27 CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
DB 657 FKPLWR 662

RESULT 6
Q9VY94
ID Q9VY94 PRELIMINARY: PRT: 917 AA.
AC Q9VY94
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE GRIP91 PROTEIN.
GN L(1)DD4 OR CG10988.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila;
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthan S., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Holt R.A., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beremand B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.K., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jafari M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Minkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Retvert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AF003493; AAF48309.1;
DR FlyBase: FBgn0001612; l(1)dd4.
DR InterPro: IPR000634; dehydrtse_ser_thr.
DR PROSITE: PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 917 AA; 103706 MW; 6AE88C211D256BB CRC64;

Query Match 94.3%; Score 33; DB 5; Length 917;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
DB 657 FKPLWR 662

RESULT 7
Q9XJ12
ID Q9XJ12 PRELIMINARY: PRT: 221 AA.
AC Q9XJ12
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 14, Last annotation update)
DE HYPOTHETICAL PROTEIN.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzoae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasuki T., Nagamura Y., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
clone:P0680A33.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB023482; BAA78741.1;
SQ SEQUENCE 221 AA; 24502 MW; DC24FF370660BC15 CRC64;

Query Match 88.6%; Score 31; DB 10; Length 221;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FRPXWR 6
DB 203 FRPTWR 208

RESULT 8
Q9D3F2
ID Q9D3F2 PRELIMINARY: PRT: 204 AA.
AC Q9D3F2

```

DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 5830420C20RIK PROTEIN.
GN 5830420C20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=THYMUS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyushaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK017942; BAB31014.1;
DR MGD: MGI:1923275; 5830420C20RIK.
SO SEQUENCE 204 AA; 22892 MW; 98AF03DF7FC7B066 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 204;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKXWR 6
Db 29 FRPRW 34

RESULT 9
Q9YE03 PRELIMINARY; PRT: 239 AA.
AC Q9YE03
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 24.8 KDA PROTEIN APE0766.
GN APE0766.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=93310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oyuchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).

```

```

DR EMBL: AP000060; BAA79744.1; -
DR InterPro: IPR001454; Hydrolase.
DR InterPro: IPR000150; Hypotheset_cof.
DR Pfam: PF00702; Hydrolase; 1.
DR PROSITE: PS01229; COF_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 239 AA; 24758 MW; EA30A3CDFD86B3DE CRC64;

Query Match 85.7%; Score 30; DB 1; Length 239;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKXWR 6
Db 35 FKPSWQ 110

RESULT 10
Q9D9R1 PRELIMINARY; PRT: 440 AA.
ID Q9D9R1
AC Q9D9R1
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ADULT MALE TESTIS CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:1700030N20, FULL INSERT SEQUENCE.
GN 5830420C20RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fieischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyushaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK006565; BAB24655.1;
DR MGD: MGI:1923275; 5830420C20RIK.
SO SEQUENCE 440 AA; 49566 MW; 585DA87E19879644 CRC64;

Query Match 85.7%; Score 30; DB 11; Length 440;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKXWR 6
Db 39 FRPRW 44

RESULT 11
Q31566 PRELIMINARY; PRT: 606 AA.
ID Q31566

```


AC O31566;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE YFIX PROTEIN.
 GN YFIX
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=9804033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Prescan E., Fujita C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seter S.J., Sertor P., Shin B.S., Solido B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambull R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat K., Yamamoto H., Yamane K., Yasumoto K., Yala K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis".
 RT Nature 390:249-256(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z99108; CAB12672.1; -;
 KW Complete proteome.
 SQ SEQUENCE 606 AA; 68677 MW; 4B58D058922ED0CA CRC64;

Query Match 85.7%; Score 30; DB 2; Length 606;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 :|||
 DB 570 YKPEWR 575

RESULT 12
 O52961
 ID O52961 PRELIMINARY; PRT: 610 AA.

AC O52961;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE YFIX.
 OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97101647; PubMed=8946165;
 RA Yamamoto H., Uchiyama S., Sekiguchi J.;
 RT Cloning and sequencing of a 27.8-kb nucleotide sequence of the 79
 RT genes-81 degrees region of the Bacillus subtilis genome containing
 RT the spe locus.
 RL CNA Res. 3:257-262(1996).
 DR EMBL: D85082; BAA24464.1; -;
 SQ SEQUENCE 610 AA; 69089 MW; 34094DD5CA72FF66 CRC64;

Query Match 85.7%; Score 30; DB 2; Length 610;
 Best Local Similarity 66.7%; Pred. No. 4.7e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 :|||
 DB 574 YKPEWR 579

RESULT 13

Q9NA80 PRELIMINARY; PRT: 684 AA.

AC Q9NA80;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE Y57A10A.20 PROTEIN.
 GN Y57A10A.20.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilia; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smye R.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=99069613; PubMed=9851916;
 RA none;

RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: AL117195; CAB55023.1; -;
 SQ SEQUENCE 684 AA; 79659 MW; F0E4FAF28EAF687 CRC64;

Query Match 85.7%; Score 30; DB 5; Length 684;
 Best Local Similarity 66.7%; Pred. No. 5.3e+02;
 Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPXWR 6
 :|||
 DB 672 FKPKWK 677

RESULT 14
 Q12361 PRELIMINARY; PRT: 961 AA.

AC Q12361;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE HYPOTHETICAL 110.7 KDA PROTEIN YDL035C.
 GN GPI OR D2749 OR YDL035C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ALPHA 5288C;
RA Paulin L., Saren A.M.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Paulin L., Saren A.M., Laamanen P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: 271781; CAA96454.1;
DR EMBL: 274083; CAA98593.1;
DR SGD: S0002193; GPR1.
KW Hypothetical protein.
SQ SEQUENCE 961 AA; 110708 MW; 989D857872A4209 CRC64;

Query Match 85.7%; Score 30; DB 3; Length 961;
Best Local Similarity 66.7%; Pred. No. 7.3e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKXWR 6
Db 156 FKPNWK 161

RESULT 15
O9LJ01
ID O9LJ01 PRELIMINARY; PRT; 1232 AA.
AC O9LJ01;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CRIAAD55299.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX PubMed-10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety p1,
RT TAC and BAC clones."
RL DNA Res. 7:217-221(2000).
CC -1- SIMILARITY: CONTAINS A RING-TYPE ZINC FINGER.
DR EMBL: AF000414; BAB01179.1;
DR InterPro: IPR000194; ATPase_alpha_beta.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF000097; zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00152; ATPASE_ALPHABETA; UNKNOWN_1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
KW Zinc-finger.
SQ SEQUENCE 1232 AA; 139223 MW; D3770C4A9D699207 CRC64;

Query Match 85.7%; Score 30; DB 10; Length 1232;
Best Local Similarity 66.7%; Pred. No. 9.2e+02;
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FKXWR 6
Db 902 FKPNWK 907

Search completed: February 27, 2002, 11:50:10
Job time: 989 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:36:04 : Search time 132.19 Seconds
(without alignments)
1.021 Million cell updates/sec

Title: US-09-446-109A-9

Perfect score: 35

Sequence: 1 FKXWR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/5B.COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	33	94.3	6	1	US-08-215-137-13
2	28	80.0	13	4	US-08-915-314-26
3	28	80.0	16	1	US-08-079-051-2
4	28	80.0	16	5	PCT-US94-06994-2
5	28	80.0	115	3	US-08-513-974B-351
6	28	80.0	243	2	US-08-829-110-3
7	28	80.0	304	1	US-08-118-270-35
8	28	80.0	304	5	PCT-US93-08528-35
9	28	80.0	350	2	US-08-458-970A-9
10	28	80.0	463	3	US-09-082-310-1
11	28	80.0	482	2	US-08-876-874-2
12	28	80.0	496	3	US-08-906-769-113
13	28	80.0	496	3	US-08-906-616-113
14	28	80.0	496	4	US-08-817-795-113
15	28	80.0	496	4	US-08-639-075A-113
16	28	80.0	496	4	US-09-012-431-113
17	28	80.0	496	4	US-09-012-692-113
18	28	80.0	496	4	US-08-906-613-113
19	28	80.0	496	5	PCT-US95-14442A-113
20	28	80.0	592	3	US-08-991-813-2
21	27	77.1	157	4	US-09-461-474-14
22	27	77.1	355	2	US-08-666-367B-6
23	27	77.1	355	4	US-09-143-438-6
24	27	77.1	448	4	US-09-461-474-8
25	27	77.1	474	4	US-09-461-474-10
26	27	77.1	527	3	US-08-907-229-2
27	27	77.1	566	2	US-08-666-367B-5

23 27 77.1 566 4 US-09-143-438-5 Sequence 5, Appli
24 27 77.1 1040 2 US-08-254-989-2 Sequence 2, Appli
25 27 77.1 1536 4 US-09-413-814-10 Sequence 10, Appli
26 27 77.1 2232 4 US-09-091-219-25 Sequence 25, Appli
27 27 77.1 2247 4 US-09-091-219-2 Sequence 2, Appli
28 26 74.3 10 2 US-08-704-655-20 Sequence 20, Appli
29 26 74.3 37 1 US-08-663-543B-4 Sequence 4, Appli
30 26 74.3 42 2 US-08-766-858A-27 Sequence 27, Appli
31 26 74.3 209 4 US-09-164-193-8 Sequence 8, Appli
32 26 74.3 273 2 US-08-997-080-75 Sequence 75, Appli
33 26 74.3 273 2 US-08-997-362-75 Sequence 75, Appli
34 26 74.3 273 3 US-08-873-970-75 Sequence 75, Appli
35 26 74.3 273 4 US-09-093-855-75 Sequence 75, Appli
36 26 74.3 336 4 US-09-334-601-13 Sequence 13, Appli
37 26 74.3 370 2 US-08-997-080-194 Sequence 194, App
38 26 74.3 370 4 US-09-095-855-194 Sequence 194, App
39 26 74.3 417 1 US-08-351-981-6 Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-215-137-13
Sequence 13, Application US/08215137
Patent No. 5614370
GENERAL INFORMATION:
APPLICANT: Konteatis, Zenon
APPLICANT: Siciliano, Salvatore J
APPLICANT: Springer, Martin S
TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN CSA ANTAGONISTS
FIELD OF INVENTION: AND ANONISTS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 125 Lincoln Ave., P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215,137
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bencen, Gerard H
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: 19108
TELECOMMUNICATION INFORMATION:
TELEPHONE (908) 594-3901
TELEFAX (908) 594-4720
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
UNREPEATEDNESS: single
POLYPEPTIDE TYPE: linear
ISOLATION: NO
SYNTHETIC: NO
FRAGMENT TYPE: C-terminal
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= BH
OTHER INFORMATION: /note="either the natural phenylalanine amino
OTHER INFORMATION: terminus or the Bolton-Hunter modified peptide

COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06994
FILING DATE: 20-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Tumarkin, Lisa A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: FD-2790
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 16 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: C5aR(133-148)
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..16
PCT-US94-06994-2

Query Match 80.0%; Score 28; DB 5; Length 16;
Best Local Similarity 80.0%; Pred. No. 10;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMW 5
DB 7 FKPMW 11

RESULT 5
US-08-513-974B-351
Sequence 351, Application US/08513974B
Patent No. 6114139
GENERAL INFORMATION:
APPLICANT: Hinuma, Shuji
APPLICANT: Hosoya, Masaki
APPLICANT: Fujii, Ryo
APPLICANT: Ohtaki, Tetsuya
APPLICANT: Fukusumi, Shoji
APPLICANT: Ohgi, Kazuhiro
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,
TITLE OF INVENTION: PRODUCTION, AND USE THEREOF
NUMBER OF SEQUENCES: 380
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, HIRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/513.974B
FILING DATE: 14-Sep-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP95/01599
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-093989
FILING DATE: 19-AUG-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-057186
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-007177
FILING DATE: 20-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-326611
FILING DATE: 28-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-270017
FILING DATE: 02-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236357
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-236356
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189274
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-189273
FILING DATE: 11-AUG-1994
APPLICATION NUMBER: JP 6-189272
FILING DATE: 11-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 45753
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 351:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-513-974B-351

Query Match 80.0%; Score 28; DB 3; Length 115;
Best Local Similarity 80.0%; Pred. No. 72;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMW 5
DB 4 FKPMW 8

RESULT 6
US-08-829-110-3
Sequence 3, Application US/08829110
Patent No. 5882890
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL REGULATORS OF G-PROTEIN
TITLE OF INVENTION: SIGNALING
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA

; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FALSESEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/29,110
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0259 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: THYMNOT02
; CLONE: 343504
; US-08-829-110-3

Query Match 80.0%; Score 28; DB 2; Length 243;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KPXR 6
Db 71 KPAWR 75

RESULT 7
US-08-118-270-35
; Sequence 35, Application US/08118270
; Patent No. 5508384
; GENERAL INFORMATION:
; APPLICANT: Murphy, Randall B.
; APPLICANT: Schuster, David I.
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/118,270
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; US-08-829-110-3

; REFERENCE/DOCKET NUMBER: MURPHY-2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 304 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-118-270-35

Query Match 80.0%; Score 28; DB 1; Length 304;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXW 5
Db 103 FKPIW 107

RESULT 8
PCT-US93-08528-35
; Sequence 35, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 304 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US93-08528-35

Query Match 80.0%; Score 28; DB 5; Length 304;
Best Local Similarity 80.0%; Pred. No. 1.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPXW 5

Db 103 FKPIW 107

RESULT 9
US-08-458-970A-9
: Sequence 9, Application US/08458970A
: Patent No. 5861272
: GENERAL INFORMATION:
: APPLICANT: L.I., ET AL.
: TITLE OF INVENTION: C5a Receptor
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CECCHIA, BYRNE, BAIN, GILFILLAN,
: STREET: 6 BECKER FARM ROAD
: CITY: ROSELAND
: STATE: NEW JERSEY
: COUNTRY: USA
: ZIP: 07068
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 INCH DISKETTE
: COMPUTER: IBM PS/2
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WORD PERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/458,970A
: FILING DATE: June 2, 1995
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/09234
: FILING DATE: 16 AUG 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: MULLINS, J.G.
: REGISTRATION NUMBER: 33,073
: REFERENCE/DOCKET NUMBER: 325800-353
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 201-994-1700
: TELEFAX: 201-994-1744
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 350 amino acids
: TYPE: amino acid
: STRANDEDNESS: linear
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-458-970A-9

Query Match 80.0%; Score 28; DB 2; Length 350;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FKPIW 5
Db 139 FKPIW 143

RESULT 10
US-09-082-310-1
: Sequence 1, Application US/09082310
: Patent No. 6096526
: GENERAL INFORMATION:
: APPLICANT: Hillman, Jennifer L.
: APPLICANT: Lal, Preeti
: APPLICANT: Corley, Neil C.
: APPLICANT: Guegler, Karl J.
: APPLICANT: Yue, Henry
: TITLE OF INVENTION: HUMAN NUCLEIC ACID METHYLASES
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 PORTER DRIVE

CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/082,310
FILING DATE: Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PF-0520 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE: BRSTNOT07
CLONE: 2124957
US-09-082-310-1

Query Match 80.0%; Score 28; DB 3; Length 463;
Best Local Similarity 80.0%; Pred. No. 2.9e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 KPSWR 6
Db 294 KPSWR 297

RESULT 11
US-08-876-874-2
: Sequence 2, Application US/08876874
: Patent No. 5942495
: GENERAL INFORMATION:
: APPLICANT: Ames, Robert
: APPLICANT: Bergsma, Derk
: APPLICANT: Foley, James
: APPLICANT: Kumar, Chandrika
: APPLICANT: Sarau, Henry
: TITLE OF INVENTION: THERAPEUTIC AND SCREENING
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: RATNER & PRESTIA
: STREET: P.O. BOX 980
: CITY: VALLEY FORGE
: STATE: PA
: COUNTRY: USA
: ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,874
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/019,627
FILING DATE: 16-JUN-1997


```
; ATTORNEY/AGENT INFORMATION:
; NAME: PRESTIA, PAUL F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: P50501
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 482 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-876-874-2

Query Match      80.0%; Score 28; DB 2; Length 482;
Best Local Similarity 80.0%; Pred. No. 3e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMW 5
DB 125 FKPMW 129

RESULT 12
US-08-906-769-113
; Sequence 113, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 24-APR-1996
; CLASSIFICATION:
; NAME: Concell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-876-874-2

Query Match      80.0%; Score 28; DB 3; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMW 5
DB 358 FKPMW 362

RESULT 13
US-08-906-616-113
; Sequence 113, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Concell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-906-616-113

Query Match      80.0%; Score 28; DB 3; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FKPMW 5
DB 358 FKPMW 362

RESULT 14
US-08-817-795-113
; Sequence 113, Application US/08817795
; Patent No. 6139840
; GENERAL INFORMATION:
```

APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Frank, Glenn R.
APPLICANT: Heath, Andrew W.
APPLICANT: Yamaka, Miles Yamanaka
APPLICANT: Arfsten, Ann
APPLICANT: Dale, Beverly
APPLICANT: Sliessler, Gary
TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
NUMBER OF SEQUENCES: 119
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,795
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCY/US95/14442
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gary J. Connell
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-817-795-113

Query Match 80.0%; Score 28; DB 4; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKW 5
DB 358 FKPKW 362

RESULT 15
US-08-639-075A-113
Sequence 113, Application US/08639075A
Patent No. 6150125
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190

CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 496 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-639-075A-113

Query Match 80.0%; Score 28; DB 4; Length 496;
Best Local Similarity 80.0%; Pred. No. 3.1e+02;
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FKPKW 5
DB 358 FKPKW 362

Search completed: February 27, 2002, 11:36:04
Job time: 143 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:54:08 ; Search time 62.61 Seconds
(without alignments)
11.681 Million cell updates/sec

Title: US-09-446-109A-23
Perfect score: .30
Sequence: 1 FPXWR 5

Scoring table: HLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq	length: 0
Maximum DB seq	length: 20

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : STPREMBL_17.*
1: sp_archea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_protist.*
12: sp_virus.*
13: sp_invertebrate.*
14: sp_unclassified.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query		Length	DB	ID	Description
		Match	Score				
1	29	96.7	101	5	Q9V671	09v671	droserophila
2	29	96.7	113	10	Q39293	Q39293	brassica nap
3	29	96.7	157	5	Q9XTY5	Q9xyt5	caenorhabdi
4	29	96.7	158	4	Q9NNW3	Q9nnw3	homo sapien
5	29	96.7	158	4	Q9NNW2	Q9nnw2	homo sapien
6	29	96.7	229	8	Q37633	Q37633	thaqoletis
7	29	96.7	245	2	Q9PDE1	Q9pde1	xylella fas
8	29	96.7	273	10	Q9SUW9	Q9suw9	arabidopsis
9	29	96.7	292	2	Q9HT83	Q9ht83	pseudomonas
10	29	96.7	321	11	Q9UD67	Q9ud67	mus musculus
11	29	96.7	332	2	Q9APJ5	Q9apj5	erwinia chr
12	29	96.7	333	2	Q9L3G9	Q9l3g9	erwinia rha
13	29	96.7	334	10	Q9FTW4	Q9ftw4	oryza sativ
14	29	96.7	349	4	Q15429	Q15429	homo sapien
15	29	96.7	417	10	Q9SAAL	Q9saal	arabidopsis
16	29	96.7	466	2	Q9FRK4	Q9frk4	zymomonas m
17	29	96.7	479	2	O86742	O86742	streptomyces
18	29	96.7	497	2	O91702	O91702	pseudomonas
19	29	96.7	501	5	O16923	O16923	caenorhabdi

20	29	96.7	502	5	Q9GUT5	Q9QUI5
21	29	96.7	555	1	Q29748	caenorhabdi
22	29	96.7	581	4	Q9BRO1	archaeoglob
23	29	96.7	612	4	Q9H8Y2	Q9BRQ1 homo sapien
24	29	96.7	621	4	Q9H9Y1	Q9H8Y2 homo sapien
25	29	96.7	625	2	Q66676	Q9H9Y1 homo sapien
26	29	96.7	634	2	Q9NW19	Q66676 aquitex aeo
27	29	96.7	634	4	Q9NWX5	Q9NW19 homo sapien
28	29	96.7	642	2	Q25396	Q9NWX5 homo sapien
29	29	96.7	642	2	Q9ZLF3	Q25396 helicobacte
30	29	96.7	650	5	Q17866	Q9ZLF3 helicobacte
31	29	96.7	735	2	Q9HU63	Q17866 caenorhabdi
32	29	96.7	735	2	Q9HU63	Q9HU63 pseudomonas
33	29	96.7	778	5	Q9U9K6	Q9HU63 pseudomonas
34	29	96.7	811	4	Q9BWX2	Q9U9K6 caenorhabdi
35	29	96.7	1071	11	Q9D2K4	Q9BWX2 homo sapien
36	29	96.7	1108	13	Q9PWD0	Q9D2K4 mus musculus
37	29	96.7	1117	5	Q9U9K7	Q9PWD0 tetraodon f
38	29	96.7	1451	2	Q45336	Q9U9K7 caenorhabdi
39	29	96.7	1451	2	Q45044	Q45336 bordetella
40	28	96.7	1464	2	Q9S5D5	Q45044 bordetella
41	28	93.3	68	2	P73858	Q9S5D5 bordetella
42	28	93.3	84	2	P74471	P73858 synecchocyst
43	28	93.3	84	2	Q35568	P74471 synecchocyst
44	28	93.3	93	2	P73028	Q35568 synecchocyst
45	28	93.3	103	10	Q9M045	P73028 synecchocyst
46	28	93.3	110	6	Q9GLJ9	Q9M045 arabidopsis
47	28	93.3	110	6	Q9GLJ9	Q9GLJ9 canis famil

ALIGNMENT'S

RESULTS

0967; AC Q5V671 PRELIMINARY; PRT; 101 AA.
AD Q5V671;
AE C.-RAY-2000 (TREMBLrel. 13, Created)
AF L.-NAY-2000 (TREMBLrel. 13, Last sequence update)
AG J.-JIT-2001 (TREMBLrel. 17, Last annotation update)
AH G3I174 PROTEIN.
AI G3I174
AJ Drosophila melanogaster (Fruit fly).
AK Charyota; Metazoa; Arthropoda; Tracheata; Insecta;
AL Phryganea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
AM Spandroidae; Drosophilidae; Drosophila.
AN W3I_TaxID=727;
AO SOURCE FROM N.A.
AP TRAIN=BERKELEY;
AQ MEDLINE=20195006; Pubmed=107311132;
AR Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
AS Avarnates P.G., Scherer S.E., Li P.W., Hoskins K.A., Galie R.F.,
AT George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
AU Jackson G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
AV Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
AW Min K.L., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
AX April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
AY Bailey R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
AZ Keelson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
BA Parkova D., Botchan M.R., Bouck J., Brocktein P., Brotlier P.,
BB Bartis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
BC Barry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
BD De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
BE Johnson K., Jupp L.E., Downes M., Nugan-Rocha S., Dunkov R.C., Dunn P.,
BF Martin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
BG Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
BH Block A., Gong F., Gorrell J.H., Gu Z., Guan P., Harlis M.,
BI Harlis N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
BJ Johnston D., Houston K.A., Howland T.J., Wei M.-H., Ihigwam C.,
BK Kalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
BL Khmaladze B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
BM Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
BN Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
BO Markulov G., Milshina N.V., Moharry C., Morris J., Moshpri A.

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Ruskern D.K., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klanos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003823; AAF58560.1;
 DR FlyBase; FBgn0033694; CG13174.
 SQ SEQUENCE 101 AA; 11820 MW; D5C4D85B514CB9B9B CRC64;

Query Match 96.7%; Score 29; DB 5; Length 101;
 Best Local Similarity 80.0%; Pred. No. 1.2e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 II II
 DB 75 FPXWR 79

RESULT 2
 Q39293
 ID Q39293 PRELIMINARY; PRT; 113 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ELONGATION FACTOR EF-1A (FRAGMENT).
 OS Brassica napus (Rape).
 OC Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN III

RP SEQUENCE FROM N.A.
 KC STRAIN-SAMOUAI; TISSUE=ETIOLATED SEEDLINGS, (5-DAYS-OLD);
 RA Saez-Vasquez J.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO GTP-BINDING ELONGATION FACTOR FAMILY.
 DR EMBL; U21744; AAA86366.1;
 DR Mendei; 15767; Hraa; 1139; 15767.
 DR InterPro: IPR000795; GTP_EFTU.
 DR Pfam: PF00009; GTP_EFTU; 1.
 KW Elongation factor; GTP-binding; Protein biosynthesis.
 FT NON_TER 1
 SQ SEQUENCE 113 AA; 12560 MW; 1497830707223948 CRC64;

Query Match 96.7%; Score 29; DB 10; Length 113;
 Best Local Similarity 80.0%; Pred. No. 1.4e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 II II
 DB 103 FPTWR 107

RESULT 3
 Q9XTY5
 ID Q9XTY5 PRELIMINARY; PRT; 157 AA.

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE T0308.2 PROTEIN.

GN T0308.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabdilloidea;
 OC Rhabdillidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN III
 RP SEQUENCE FROM N.A.
 RA Mortimore B.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 EX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Cooper J., Coulson A.,
 RA Holford J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lichting J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thertory-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Wellstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";
 RL Nature 368:32-38(1994).
 DR EMBL; Z92838; CAB07406.1;
 DR InterPro: IPR000230; Ribosomal_S12.
 DR Pfam: PF00164; Ribosomal_S12; 1.
 DR PRINTS; PR01034; RIBOSOMALS12.
 DR ProDom: PD000576; Ribosomal_S12; 1.
 SQ SEQUENCE 157 AA; 17370 MW; 6C6A837282EB26C CRC64;

Query Match 96.7%; Score 29; DB 5; Length 157;
 Best Local Similarity 80.0%; Pred. No. 1.9e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 FPXWR 5
 II II
 DB 40 FPAWR 44

RESULT 4
 Q9NNW3

ID Q9NNW3 PRELIMINARY; PRT; 158 AA.
 AC Q9NNW3;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE MEMBRANE COFACTOR PROTEIN CD46 VARIANT (FRAGMENT).
 GN MCP.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN III
 RP SEQUENCE FROM N.A.
 EX MEDLINE=20227770; PubMed=10751138;
 RA Kusuhabara K., Sasaki Y., Nakao F., Ihara K., Hattori H., Yamashita S.,
 RA Nishi K., Koide N., Aiba H., Takeshita K., Hara T.;
 RT "Analysis of measles virus binding sites of the CD46 gene in patients
 with subacute sclerosing panencephalitis.";
 RL J. Infect. Dis. 181:1447-1449(2000).
 DR EMBL; AF209713; AAF73845.1;
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR Pfam: PF00081; sushi; 2.
 DR SMART; SM00072; CCP; 2.
 FT VARIANT 59 59 Q -> R.
 FT NON_TER 158 158
 SQ SEQUENCE 158 AA; 18098 MW; 4E3F07EC5C454F5 CRC64;

Query Match 96.7%; Score 29; DB 4; Length 158;

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zeldanis J., Zetubal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 406:151-159(2000).
 DR EMBL: AE003974; AAF84247.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 245 AA; 27193 MW; 16C8ADB9FC490455 CRC64;

Query Match: 96.7%; Score 29; DB 2; Length 245;

Best Local Similarity 80.0%; Pred. No. 2.8e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5

Db 197 FPSWR 201

RESULT 8

Q9SU99 PRELIMINARY; PRT: 273 AA.

AC Q9SU99;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE HYPOTHETICAL 31.2 KDA PROTEIN.

GN FK2.180 OR AT4G22600.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Wedler H., Wambutt R., Bancroft I., Mewes H.W.,

RA Mayer K.F.X., Lemcke K., Schueller C.

RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL033545; CAB22165.1;

DR EMBL: AL161557; CAB79215.1;

KW Hypothetical protein.

SQ SEQUENCE 273 AA; 31248 MW; 7B564F686CE8EB90 CRC64;

Query Match 96.7%; Score 29; DB 10; Length 273;

Best Local Similarity 80.0%; Pred. No. 3.1e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5

Db 79 FPSWR 83

RESULT 9

Q9HT83 PRELIMINARY; PRT: 292 AA.

AC Q9HT83;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE HYPOTHETICAL PROTEIN PA5488.

GN PA5488.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brady L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an

opportunistic pathogen,"

KL Nature 406:959-964(2000).

DR EMBL: AE004961; AAG08873.1;

DR InterPro; IPR000504; RRM.

DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.

EW Hypothetical protein; Complete proteome.

SQ SEQUENCE 292 AA; 32949 MW; 14610337E8CCDFD4 CRC64;

Query Match 96.7%; Score 29; DB 2; Length 292;

Best Local Similarity 80.0%; Pred. No. 3.3e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 FPXWR 5

Db 231 FPSWR 235

RESULT 10

Q9D6L7

ID C9J6L7 PRELIMINARY; PRT: 321 AA.

AC Q9J6L7;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE 23100760141K PROTEIN.

GN 23100760141K.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=TONGUE;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakura I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Adota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Flutschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sasaki K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Hernaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsaki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection,"

KL Nature 409:685-690(2001).

DR EMBL: AK010208; BAB26769.1;

DR MGD: MGI:1919189; 2310076014Rik.
 DR InterPro: IPR000847; HTH_LysR.
 DR PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN_1.
 SQ SEQUENCE 321 AA; 35038 MW; 6B27EBC746552DFB CRC64;

Query Match 96.7%; Score 29; DB 11; Length 321;
 Best Local Similarity 80.0%; Pred. No. 3.7e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
 || ||
 Db 44 FPSWR 48

RESULT 11

ID O9APJ5 PRELIMINARY; PRT; 332 AA.
 AC O9APJ5.
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ENDO-1,4-BETA-D-GLUCANASE PRECURSOR.
 GN CEL8Y.
 OS Erwinia chrysanthemi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=556;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PV35;
 RA Cho S.J., Park S.R., Yun H.D.;
 RT "Cloning and sequencing of a cel8Y gene of Pectobacterium chrysanthemi".
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 KM EMBL: AF282321; AAC49556.1;
 KW Signal.
 FT SIGNAL 1 23 POTENTIAL.
 SQ SEQUENCE 332 AA; 37627 MW; A1B5D7B0CBH20RE6 CRC64;

Query Match 96.7%; Score 29; DB 2; Length 332;
 Best Local Similarity 80.0%; Pred. No. 3.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
 || ||
 Db 178 FPAWR 182

RESULT 12

ID O9L3G9 PRELIMINARY; PRT; 333 AA.
 AC O9L3G9.
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ENDOGLUCANASE PRECURSOR (NC 3.2.1.4).
 GN CELA.
 OS Erwinia raphanotici.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Erwinia.
 OX NCBI_TaxID=55212;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Saarliahli H.T.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCPPB2989;
 RA Riekkilä R.;
 RT "Members of the amylovora group of Erwinia are cellulolytic and

KT possess genes homologous to the type II secretion pathway."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ276358; CAB89803.1;
 DR InterPro: IPR002037; Glyco_hydro_8.
 DR Pfam: PF01270; Glyco_hydro_8; 1.
 DR PRINTS: PR00735; GLHYDRASE8.
 DR Signal; Hydrolase; Glycosidase.
 ZW SIGNAL 1 23 POTENTIAL.
 FT SIGNAL 1 23 ENDOGLUCANASE.
 FT CHAIN 24 333
 SQ SEQUENCE 333 AA; 37783 MW; D75CFF212302673A CRC64;

Query Match 96.7%; Score 29; DB 2; Length 333;
 Best Local Similarity 80.0%; Pred. No. 3.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
 || ||
 Db 178 FPAWR 182

RESULT 13

ID O9FTM4 PRELIMINARY; PRT; 334 AA.
 AC O9FTM4.
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE P0005A05.25 PROTEIN.
 GN P0005A05.25.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(CA3) genomic DNA, chromosome 1, PAC
 clone:P0005A05."
 ZL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP002863; BAB16921.1;
 DR InterPro: IPR000520; Exonuclease.
 DR Pfam: PF00929; Exonuclease; 1.
 DR Pfam: PF00096; zf-C2H2; 1.
 DR SMART: SM00479; EXOIII; 1.
 DR SMART: SM00355; ZnF_C2H2; 2.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 DR PROSITE: PS0157; ZINC_FINGER_C2H2_2; 1.
 KW DNA-binding; Zinc-finger.
 SQ SEQUENCE 334 AA; 37172 MW; A34531D625704A88 CRC64;

Query Match 96.7%; Score 29; DB 10; Length 334;
 Best Local Similarity 80.0%; Pred. No. 3.8e+02;
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 FPXWR 5
 || ||
 Db 101 FPAWR 305

RESULT 14

ID Q15429 PRELIMINARY; PRT; 349 AA.
 AC Q15429.
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CD46.
 OS Homo sapiens (human).

Search completed: February 27, 2002, 11:54:08
Job time: 228 sec

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hara T., Suzuki Y., Nagasawa S., Seya T.;
RT "cDNA cloning and characterization of human sperm CD46."
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: D84105; BAA12224.1; -.
DR HSSP: P10998; 1VVD.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR Pfam: PF00084; sushi; 4.
DR SMART: SM00032; CCP; 4.
SQ SEQUENCE 349 AA: 39325 MW: 8EFCDA30D3C818E CRC64;

Query Match 96.7%; Score 29; DB 4; Length 349;
Best Local Similarity 80.0%; Pred. No. 4e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
II II
DB 11 FPSWR 15

RESULT 15

Q9SAA1 PRELIMINARY; PRT: 417 AA.
AC Q9SAA1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE F25C20.9.
GN F25C20.9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Lee J.N., Kremenetskaia I., Lueros J., Ngan L., Liu A.,
RA Gonzalez A., Altafi H., Araujo R., Chao O., Conn L., Conway A.B.,
RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT "Arabidopsis thaliana chromosome 1 RAC F25C20 sequence."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007296; AAD30247.1; -.
SQ SEQUENCE 417 AA: 45985 MW: 866EBF06AC19A63D CRC64;

Query Match 96.7%; Score 29; DB 10; Length 417;
Best Local Similarity 80.0%; Pred. No. 4.7e+02;

Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 FPXWR 5
II II
DB 377 FPAWR 381

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:51:50 ; Search time 66.26 Seconds
(without alignments)
7.825 Million cell updates/sec

Title: US-09-446-109A-24
Perfect score: 36
Sequence: 1 KFXPXR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq.1101.*
1: /SIDSB/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSB/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSB/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSB/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSB/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSB/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSB/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSB/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSB/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSB/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSB/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSB/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSB/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSB/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSB/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSB/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSB/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSB/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSB/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSB/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSB/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSB/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	88.9	39	20	AAW74031 Human D2H binding
2	32	88.9	392	21	AAAG21559 Arabidopsis thalia
3	32	88.9	399	21	AAAG21558 Arabidopsis thalia
4	32	88.9	588	21	AAAG47006 Arabidopsis thalia
5	32	88.9	594	21	AAAG47005 Arabidopsis thalia
6	29	80.6	143	21	AAAY92706 Geminitovirus Rep C2
7	29	80.6	179	21	AAAG58145 Arabidopsis thalia
8	29	80.6	179	21	AAAG60688 Arabidopsis thalia
9	29	80.6	179	21	AAAG60692 Arabidopsis thalia
10	29	80.6	287	21	AAAG59876 Arabidopsis thalia
11	29	80.6	335	21	AAAY92318 Mastrevirus Rep in

32	29	80.6	342	21	AAAG60687 Arabidopsis thalia
33	29	80.6	342	21	AAAG60691 Arabidopsis thalia
14	29	80.6	343	21	AAAG58144 Arabidopsis thalia
15	29	80.6	354	20	AAAY37369 Protein involved i
16	28	77.8	71	22	AAAM16826 Peptide #3260 enco
17	28	77.8	71	22	AAAM29310 Peptide #3347 enco
18	28	77.8	71	22	AAAM4538 Peptide #3220 enco
19	28	77.8	82	21	AAAY3060 Human secreted pro
20	28	77.8	105	21	AAAB18079 Eucalyptus grandis
21	28	77.8	208	20	AAAW89232 Human osteoleprotege
22	28	77.8	272	17	AAAR99944 Mutated OCIF, OCIF
23	28	77.8	277	21	AAAB41139 Human ORFX ORF903
24	28	77.8	321	17	AAAR99949 Mutated OCIF, OCIF
25	28	77.8	327	17	AAAR99941 Mutated OCIF, OCIF
26	28	77.8	349	20	AAAW83928 Human FTHMA-070 pa
27	28	77.8	351	17	AAAR99943 Mutated OCIF, OCIF
28	28	77.8	355	22	AAAB61888 Soybean HES1 homol
29	28	77.8	357	22	AAAM24361 Human EST encoded
30	28	77.8	358	22	AAAB61886 Soybean HPS1 homol
31	28	77.8	359	17	AAAR99937 Mutated OCIF, OCIF
32	28	77.8	359	17	AAAR99939 Mutated OCIF, OCIF
33	28	77.8	360	17	AAAR99936 Mutated OCIF, OCIF
34	28	77.8	360	17	AAAR99938 Mutated OCIF, OCIF
35	28	77.8	380	17	AAAR99924 Mature osteoclasto
36	28	77.8	380	22	AAAB60988 Murine ORC cystein
37	28	77.8	383	22	AAU00428 Rat Gas1 protein.
38	28	77.8	390	17	AAAR99357 Human tumour necro
39	28	77.8	391	19	AAAW53238 Human OCIF genome
40	28	77.8	391	21	AAAG16554 Arabidopsis thalia
41	28	77.8	391	21	AAAG35221 Zea mays protein f
42	28	77.8	393	17	AAAR99948 Mutated OCIF, OCIF
43	28	77.8	395	19	AAAW5636 Modified TR1 recep
44	28	77.8	395	21	AAAB18716 Carboxy terminus m
45	28	77.8	399	17	AAAR99942 Mutated OCIF, OCIF

ALIGNMENT

RESULTS 1
AD AAW74031 standard; peptide; 39 AA.
XX
XX AAW74031;
XX
XX 10-MAY-1999 (first entry)
XX
XX Human D2H binding protein DAB3.
XX
XX Q5A10-intestinal transport receptor; binding protein; hSI; HPT1;
XX Q5A10; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
XX Intestinal peptide-associated transporter; hypertension; diabetes;
XX Q5A10porosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
XX Q5A10; therapeutic agent delivery; therapy.
XX
XX Homo sapiens.
XX
XX W69A1325-A2.
XX
XX 10-MAY-1998 98WO-US10088.
XX
XX 10-MAY-1997 97US-0046595.
XX
XX (CYP2-) CYTOSOL CORP.
XX (ELAN-) ELAN CORP PLC.
XX Alvarez VL, Bellinka BA, Cogney GM, Carter JM, Lambkin IJ;
XX McMahon DJ, Patterson CA, Singleton J;
XX WPT: 1999-009568/01.

PT New proteins that bind specifically to receptors in the
PT gastro-intestinal tract and related nucleic acid - chimaeras and
PT antibodies, used to deliver therapeutic or diagnostic agents to, or
PT through, the gastrointestinal tract, e.g. insulin or leuprolide
XX PS
XX Claim 2: Page 54: 294pp: English.
XX
CC This sequence represents a peptide that specifically binds to the human
CC D2H protein. The invention relates to purified proteins (I) that bind
CC specifically to at least one of the gastro-intestinal (GI) tract
CC receptors human intestinal peptide-associated transporter (HPI),
CC hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
CC active transport of therapeutic agents through human and animal GI tissue
CC (into the blood) for in vivo delivery, particularly for treatment or
CC prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
CC cancer, migraine, or angina pectoris. Specifically they are used to
CC deliver insulin or leuprolide, but many other suitable therapeutic agents
CC are disclosed, including genes or inhibitory nucleic acid, imaging agents
CC and antigens. (I) may also provide targeting to the GI tract. Other uses
CC of (I) are: (i) to determine the level of specified receptors in a sample
CC (in a binding assay); and (ii) to screen for molecules that bind (I).
CC immunogenic analogues or derivatives of (I) are used to raise antibodies
CC and in immunoassays. The antibodies are used to locate, detect and
CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC etc., also for peptide purification and immobilisation.
XX
XX Sequence 39 AA:

Query Match 88.9%; Score 32; DB 20; Length 39;
Best Local Similarity 71.4%; Pred. No. 7;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KFXPXWR 7
Db 14 KFXPWR 20

RESULT 2

AAG21559 ID AAG21559 standard; Protein: 392 AA.

XX AC AAG21559;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 24159.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 15-JUL-1999; 99US-0144086.
PR 15-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.

```
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 23-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.

PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 88.9% Score 32; DB 21; Length 392;
Best Local Similarity 71.4%; Pred No. 71;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPKWR 7
DD 118 kfgpkwr 124

RESULT 3
AAAG21558
ID AAAG21558 standard; Protein; 399 AA.
XX
AC AAAG21558;
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24158.
SW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
TN EF-U33405-A2.
XX
PI 06-SEP-2000.
XX
FF 15-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 09-APR-1999; 99US-0128714.
PR 15-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132486.
```

PK 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0132487.
PR 23-JUL-1999; 99US-0132863.
PR 23-JUL-1999; 99US-0134256.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 26-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 19-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

```

PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      88.9%; Score 32; DB 21; Length 399;
Rest Local Similarity 71.4%; Pred. No. 72;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KEXPXWR 7
DB 125 kfgpkwr 131

RESULT 4
AAG47006
ID AAG47006 standard; Protein; 588 AA.
XX
AC AAG47006;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 59199.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
IN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0121180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

```

```
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 28-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151130.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.

Query Match 88.9%; Score 32; DB 21; Length 588;
Best Local Similarity 71.4%; Pred. NO. 1.1e+02;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KEXPXWR 7
Db 118 kfgpkwr 124

RESULT 5
AAG47005
ID AAG47005 standard; Protein: 594 AA.
XX AC AAG47005;
XX DT 18-OCT-2000 (First entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 59198.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PE 25-FEB-2000; 2000EP-0301439.
XX PF 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
```


PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 05-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 07-AUG-1999; 99US-0147493.
PR 08-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 16-OCT-1999; 99US-0158369.
PR 17-OCT-1999; 99US-0159293.
PR 18-OCT-1999; 99US-0159294.
PR 18-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 23-OCT-1999; 99US-0160980.
PR 23-OCT-1999; 99US-0160981.
PR 23-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 88.9%: Score 32; DB 21; Length 594;
 Best Local Similarity 71.4%: Pred. No. 1.le+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPXR 7
 II I I I
 Db 124 kfgpkwr 130

RESULT 6

AA92706
 ID AA92706..Standard; Protein; 143 AA.

XX
 AC AA92706;

DT 10-AUG-2000 (first entry)

DE Geminivirus Rep C2 product.

XX Rep; replicase; rescue; replication; vector; gene expression;
 KW recombinant protein production; plant.

XX Geminivirus.

XX WO200020557-A2.

PN 13-APR-2000.

PD 07-OCT-1999; 99WO-US23520.

PF 07-OCT-1998; 98US-010352.

PR (UYCO-) UNIV CORNELL BOYCE INST PLANT RES.

PA Mason HS, Palmer KE, Mor TS, Hefferon KL, Arntzen C;

PI WPI: 2000-303756/26.

DR N-PSDB; AAA09238.

XX Nucleic acids comprising a portion of a long intergenic region of a
 PT geminivirus genome and a sequence comprising a geminiviral replicase
 PT gene, useful for amplifying a gene of interest and overproducing a
 PT protein of interest in plants

XX Disclosure: Fig 10; 121pp; English.

XX The geminivirus rep gene sequence contains a C1 open reading frame (ORF)
 CC which encodes RepA. Rep is the translation product of the spliced C1 and
 CC C2 ORFs. Upon transcription and translation, the expressed protein
 CC replicase can act in trans to effect rescue and replication of a desired
 CC heterologous nucleotide sequence. A pair of nucleic acids comprising a
 CC portion of a long intergenic region (link) of a geminivirus genome lacking
 CC a sequence encoding a geminiviral coat protein, and a nucleic acid
 CC comprising a geminiviral replicase gene linked to a fruit ripening
 CC dependent promoter, are new. Strains of *Escherichia coli* and
 CC *Agrobacterium tumefaciens* transfected with an expression vector
 CC containing the nucleic acids and a gene of interest are useful for the
 CC production of transgenic plants (or cells). The sequences and methods
 CC are useful for amplifying a gene of interest and overproducing a protein
 CC of interest in recombinant plants.

XX Sequence 143 AA;

Query Match 80.6%: Score 29; DB 21; Length 143;

Best Local Similarity 57.1%: Pred. No. 1e+02;

Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPXR 7
 II I I I
 Db 65 kfvplwk 71

RESULT 7

AA958145
 ID AA958145 standard; Protein; 179 AA.

XX
 AC AA958145;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 75022.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 01-APR-1999; 99US-0126785.

XX 06-APR-1999; 99US-0127462.

XX 08-APR-1999; 99US-0128234.

XX 16-APR-1999; 99US-0128714.

XX 19-APR-1999; 99US-0129845.

XX 21-APR-1999; 99US-0130077.

XX 23-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 26-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134768.

XX 20-MAY-1999; 99US-0134941.

XX 21-MAY-1999; 99US-0135124.

XX 24-MAY-1999; 99US-0135353.

XX 25-MAY-1999; 99US-0135629.

XX 27-MAY-1999; 99US-0136021.

XX 28-MAY-1999; 99US-0136392.

XX 31-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 27-JUN-1999; 99US-0137724.

XX 09-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 16-JUN-1999; 99US-0139453.

XX 17-JUN-1999; 99US-0139492.

XX 18-JUN-1999; 99US-0139454.

XX 18-JUN-1999; 99US-0139455.

XX 18-JUN-1999; 99US-0139456.

XX 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 06-JUL-1999; 99US-0142055.
 PR 08-JUL-1999; 99US-0142390.
 PR 09-JUL-1999; 99US-0142803.
 PR 12-JUL-1999; 99US-0142920.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 19-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 22-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 23-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148505.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 22-AUG-1999; 99US-0149902.
 PR 22-AUG-1999; 99US-0149930.
 PR 22-AUG-1999; 99US-0150566.
 PR 22-AUG-1999; 99US-0150884.
 PR 22-AUG-1999; 99US-0151065.
 PR 22-AUG-1999; 99US-0151066.
 PR 22-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 18-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 03-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 12-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 23-OCT-1999; 99US-0161404.
 PR 23-OCT-1999; 99US-0161405.
 PR 23-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 29-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 80.6%; Score 29; DB 21; Length 179;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Caps 0;

QY : KFXPW 6
 ||||
 Qb 70 kfspsw 75

RESULT 8
 ANG506a8
 ID ANG50688 standard; Protein; 179 AA.
 XZ
 AC ANG50688;
 XX

DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 78638.
DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS EP1033405-A2.
PN 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 26-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 19-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

```

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151103.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158232.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      80.6%; Score 29; DB 21; Length 179;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KFXPMW 6
Db 70 kfspaw 75

RESULT 9
AAG60692
ID AAG60692 standard; Protein: 179 AA.
XX
AC AAG60692;
XX
DY 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 78644.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

```

```

XX
OS Arabidopsis thaliana.
XX
PN EP1013405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 25-MAR-1999; 99US-0123180.
PR 29-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 25-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 31-JUN-1999; 99US-0137222.
PR 33-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 09-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 15-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 20-JUN-1999; 99US-0139492.
PR 23-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.

```

```
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149420.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151309.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 19-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match Similarity 80.6%; Score 29; DB 21; Length 179;
Best local Similarity 86.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXW 6
   |||
Db 70 kfspaw 75

RESULT 10
AAG59876
ID AAG59876 standard; Protein; 287 AA.
AX
AC AAG59876;
XX
XX
XX 18-OCT-2000 (first entry)
UE Arabidopsis thaliana protein fragment SEQ ID NO: 77496.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP:033405-A2.
XX
PD 06-SEP-2000.
XX
```

PF 25-FEB-2000; 2000EP-0301439.
XX 99US-0121825.
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 26-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 23-JUL-1999; 99US-0144352.
PR 30-JUL-1999; 99US-0144632.
PR 30-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 25-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 06-AUG-1999; 99US-0147935.
PR 06-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 19-AUG-1999; 99US-0149426.
PR 19-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 23-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 31-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 15-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

```

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      80.6%; Score 29; DB 21; Length 287;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KFXPXW 6
Db 233 kispaw 238

RESULT 11
AAY92318
XX AAY92318 standard; Protein: 335 AA.
XX AC
XX AAY92318;
XX 10-AUG-2000 (first entry)
XX Mastrevirus Rep intronless gene sequence product.
XX Rep: replicase; rescue: replication; vector: gene expression;
XX recombinant protein production; plant; intronless.
XX Mastrevirus bean yellow dwarf virus.
XX WO200020557-A2.
XX 13-APR-2000.
XX 07-OCT-1999; 99WO-US23520.
XX 07-OCT-1998; 98US-0103352.
XX (UYCÜ-) UNIV CORNELI. HOYCE INST PLANT RES.
XX Mason IIS, Palmer KE, Mor TS, Heiteron KL, Arntzen C;
PI Mason IIS, Palmer KE, Mor TS, Heiteron KL, Arntzen C;
XX Mason IIS, Palmer KE, Mor TS, Heiteron KL, Arntzen C;

DR WPT: 2000-303756/26.
DR N-PSDB; ANA09239.
XX Nucleic acids comprising a portion of a long intergenic region of a
XX geminivirus genome and a sequence comprising a geminiviral replicase
XX gene, useful for amplifying a gene of interest and overproducing a
XX protein of interest in plants
XX Disclosure; Fig 18; 121pp; English.
XX This is the product of an intronless rep gene sequence from Mastrevirus
XX for use in the invention. Removal of the intron greatly enhances
XX bean yellow dwarf virus replication. Upon transcription and translation,
XX the expressed protein replicase can act in trans to effect rescue and
XX replication of a desired heterologous nucleotide sequence. A pair of
XX nucleic acids comprising a portion of a long intergenic region (LIR) of a
XX geminivirus genome lacking a sequence encoding a geminiviral coat
XX protein, and a nucleic acid comprising a geminiviral replicase gene
XX linked to a fruit ripening-dependent promoter, are new. Strains of
XX Escherichia coli and Agrobacterium tumefaciens transfected with an
XX expression vector containing the nucleic acids and a gene of interest are
XX useful for the production of transgenic plants (or cells). The sequences
XX and methods are useful for amplifying a gene of interest and
XX overproducing a protein of interest in recombinant plants.
XX Sequence 335 AA;

Query Match      80.6%; Score 29; DB 21; Length 335;
Best Local Similarity 57.1%; Pred. No. 2.5e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KFXPXW 7
Db 257 kfvp1wk 263

RESULT 12
AAG60687
XX AAG60687 standard; Protein: 342 AA.
XX AC
XX AAG60687;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SKQ ID NO: 78637.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EF1033405-A2.
XX 05-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 25-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.

```


PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135144.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151086.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 31-SEP-1999; 99US-0151930.
PR 37-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.6%; Score 20; DB 21; Length 342;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPW 6
II I I
Db 233 kispaw 238

RESULT 13
AAG60691
ID AAG60691 standard; Protein: 342 AA.

AC AAG60691;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 78643.

DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-Sep-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132484.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 15-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.

```
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.6%; Score 29; DH 21; Length 342;
Best local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY 1 KFXPXW 6
DB 233 kfs paw 238

RESULT 14
AAG58144
ID AAG58144 standard; Protein: 343 AA.
XX
AC AAG58144;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 75021.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
CJ 03-SEP-2000.
XX
EF 23-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 03-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 27-MAR-1999; 99US-0126785.
PR 31-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 13-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 03-MAY-1999; 99US-0132407.
PR 03-MAY-1999; 99US-0132484.
PR 03-MAY-1999; 99US-0132485.
PR 03-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 12-MAY-1999; 99US-0134768.
```

PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 13-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151920.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.

PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 80.6%; Score 29; DH 21; Length 343;
 Best Local Similarity 66.7%; Pred. No. 2.5e+02;
 Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXEXW 6
 II.I.I.
 Db 234 KFSQW 239

RESULT 15

AAV37369
 ID AAV37369 standard; Protein; 354 AA.

XX AC AAV37369;

XX DT 07-OCT-1999 (first entry)

XX DE Protein involved in intermediate metabolism of nucleic acids.

XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; peritropalitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX OS Chlamydia trachomatis.

XX PN W0928475-A2.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98WO-1B01939.

XX PR 04-NOV-1998; 98US-0107077.

XX PR 28-NOV-1997; 97ER-0015041.

XX PR 17-DEC-1997; 97PR-0016034.

XX PA (CEST) GENSET.

XX PI Griffiths R;

XX DR WPI; 1999-371125/31.

XX PS Genome sequence of Chlamydia trachomatis

XX PS Disclosure; Page 1086; 1755pp; English.

CC CC AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Anticapsule and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC peritropalitis, bartholinitis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX SQ Sequence 354 AA;

Query Match

Best Local Similarity 80.6%; Score 29; DB 20; Length 354;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXEXW 7

Db 4; KFIPIWK 49

Search completed: February 27, 2002, 11:51:51
 Job time 628 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:52:58 ; Search time 34.86 Seconds
(without alignments)
15,296 Million cell updates/sec

Title: us-09-446-109a-24

Perfect score: 36
Sequence: 1 KFXPXWR 7

Scoring table: BIOSUM62
Gapop 10.0 , Capext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR-68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	94.4	639	1 WMV070	70K protein - pota
2	33	91.7	639	2 S03547	hypothetical prote
3	32	88.9	380	2 C64364	formate hydrogenly
4	32	88.9	462	2 T01549	hypothetical prote
5	32	88.9	591	2 T10640	hypothetical prote
6	31	86.1	254	2 B72549	hypothetical prote
7	30	83.3	317	2 T00550	probable phospholi
8	30	83.3	351	2 S49387	replication-associ
9	30	83.3	351	2 S42356	replication-associ
10	30	83.3	389	2 G83317	conserved hypotet
11	30	83.3	1683	2 S38103	SP014 protein - ye
12	29	80.6	85	2 T39205	dyein light chain
13	29	80.6	122	2 C81198	hypothetical prote
14	29	80.6	122	2 C81775	hypothetical prote
15	29	80.6	136	2 J01359	C2 protein - Misco
16	29	80.6	146	2 C42452	C2 protein - tobac
17	29	80.6	159	2 C85567	hypothetical prote
18	29	80.6	340	2 A86000	hypothetical prote
19	29	80.6	340	2 A67129	hypothetical 38.5
20	29	80.6	342	2 A71474	probable sulfite s
21	29	80.6	354	2 T39599	conserved hypotet
22	29	80.6	378	1 S01190	ubiquinol-cytochr
23	29	80.6	382	2 S47444	probable integrase
24	29	80.6	422	2 C82666	conserved hypotet
25	29	80.6	477	2 S77373	hypothetical prote
26	29	80.6	512	2 T48462	cytochrome p450-li
27	29	80.6	542	2 C84911	probable anion exc
28	29	80.6	563	2 S78224	virulence-associat
29	29	80.6	589	2 T18239	transcription effe

ALIGNMENTS

RESULT: 1

WMV070

70K protein - potato leaf roll virus (strain 1)

C:Species: potato leaf roll virus

A:Note: host Solanum tuberosum (potato)

C:Data: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 21-Jul-2000

C:Accession: JA0118; S24591

R:Mayo, M.A.; Robinson, D.J.; Jolly, C.A.; Hyman, L.

J. Gen. Virol. 70, 1037-1051, 1989

A:Title: Nucleotide sequence of potato leafroll luteovirus RNA.

A:Reference number: JA0119; MUID:89279282

A:Accession: JA0118

A:Molecule type: genomic RNA

A:Residues: 1-639 <MAY>

A:Cross-references: EMBL:X14600; MID:g222293; PIDN:BA00417.1; PID:g222297

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, March 1989

C:Comment: The genome is a single-stranded, positive-sense RNA.

C:Superfamily: potato leaf roll virus 70K protein

Query Match 94.4%; Score 34; DB 1; Length 639;
Best Local Similarity 71.4%; Pred. No. 24;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY : KFXPXWR 7

||| ||

Lb 614 KFTPSWR 520

RESULT: 2

S03547

hypothetical protein 2 - potato leaf roll virus

C:Species: potato leaf roll virus

C:Data: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 20-Sep-1999

:Accession: S03547

R:van der Wilk, P.; Huisman, M.J.; Cornelissen, B.J.C.; Huttinga, H.; Goldbach, R.

PERs Lett. 245, 51-56, 1989

A:Title: Nucleotide sequence and organization of potato leafroll virus genomic RNA.

A:Reference number: S03546; MUID:89171329

A:Accession: S03547

A:Molecule type: genomic RNA

A:Residues: 1-639 <VAN>

A:Cross-references: EMBL:Y07496; MID:g61198; PIDN:CAA68795.1; PID:g61200

C:Superfamily: potato leaf roll virus 70K protein

Query Match 91.7%; Score 33; DB 2; Length 639;
Best Local Similarity 71.4%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CY : KFXPXWR 7

||| ||

Db 614 KFPSPWR 620

RESULT 3
C64364
formate hydrogenlyase, subunit 5 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C:Accession: C64364
R:Ruit, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Hlako,
; Reich, C.J.; Overbeck, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999
A:Accession: C64364
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-360 <BUIL>
A:Cross-references: GR:U67501; GB:L77117; NID:g2826289; PIDN:AAB98504.1; PID:g1591218;
C:Genetics:
C:Superfamily: NADH dehydrogenase (ubiquinone) 49K protein

Query Match 88.9%; Score 32; DB 2; Length 380;
Best Local Similarity 71.4%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7
II I I I
DB 240 KFXPWWR 246

RESULT 4
T01549
hypothetical protein A.TM018A10.1 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01549
R:Dempsey, S.; Harper, M.
A:Description: The sequence of A. thaliana TM018A10.
A:Reference number: 214348
A:Accession: T01549
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-462 <EMBL>
A:Cross-references: EMBL:AF013294; NID:g2252848; PID:g2252849
A:Experimental source: Cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 107/1; 143/2
A:Note: A.TM018A10.1

Query Match 88.9%; Score 32; DB 2; Length 462;
Best Local Similarity 71.4%; Pred. No. 44;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7
II I I I
DB 327 KFPSPWR 333

RESULT 5
T10640
hypothetical protein T13K14.140 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999
C:Accession: T10640
R:Devan, M.; Pohl, T.; Weizenegger, T.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.; Lemcke

submitted to the Protein Sequence Database, June 1999

A:Reference number: 216991
A:Accession: T10640
A:Molecule type: DNA
A:Residues: 1-591 <BEV>
A:Cross-references: GSPDB:GN00062; ATSP:T13K14.140
A:Experimental source: Cultivar Columbia; BAC clone T13K14
C:Genetics:
A:Geno: ATSP:T13K14.140
A:Map position: 4

Query Match 88.9%; Score 32; DB 2; Length 591;
Best Local Similarity 71.4%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7
II I I I
DB 122 KFXPWWR 128

RESULT 6
B72549
hypothetical protein APE1678 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: B72549
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A:Reference number: A72450; MUID:99310339
A:Accession: B72549
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-254 <KAW>
A:Cross-references: DBJ:AP000062; NID:g5105244; PIDN:BAAR0679.1; PID:d1044465; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1678
C:Superfamily: Aeropyrum pernix hypothetical protein APE1678

Query Match 86.1%; Score 31; DB 2; Length 254;
Best Local Similarity 57.1%; Pred. No. 40;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7
II I I I
DB 168 KFXPAWK 174

RESULT 7
T00550
Probable phospholipase AL2g39400 [Imported] - Arabidopsis thaliana
A:Alternate names: lysophospholipase homolog F12L6.6
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 16-Feb-2001
C:Accession: T00550; H84816
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
submitted to the EMBL Data Library, July 1998
A:Description: Arabidopsis thaliana chromosome II BAC F12L6 genomic sequence.
A:Reference number: 214168
A:Accession: T00550
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-317 <ROU>
A:Cross-references: EMBL:AC004218; NID:g3355463; PID:g3355469
A:Experimental source: Cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
R.; Kaul, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, I.; Tallon,
euss, S.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Ventler
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Accession: A84420; MUID:20083487
A:Molecule type: DNA
A:Status: preliminary
A:Cross-references: GB:AE002093; MID:g3355469; PIDN:AAC27831.1; GSPDB:GN00139
A:Residues: 1-317 <STO>
C:Genetics:
A:Map position: 2
A:Introns: 55/1; 104/1; 147/3; 193/3; 226/3

Query Match 83.3%; Score 30; DB 2; Length 317;

Best Local Similarity 57.1%; Pred. No. 78;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7

II I I

Db 167 KFIPTWK 173

RESULT 8

S49387

replication-associated protein - wheat dwarf virus (French isolate)

N:Alternate names: CI-1/CI-2 composite protein

C:Species: wheat dwarf virus

A:Variety: French isolate

C:Date: 16-Feb-1995 #sequence_revision 06-Dec-1996 #text_change 21-Aug-1998

C:Accession: S49387

R:Bendaiane, M.; Schalk, H.J.; Gronenborn, B.

submitted to the EMBL Data Library, October 1994

A:Description: Identification and characterization of wheat dwarf virus (WDV) from France

A:Reference number: S49385

A:Accession: S49387

A:Molecule type: DNA

A:Residues: 1-351 <GEN>

A:Cross-references: EMBL:X82104

A:Experimental source: French isolate

A:Note: ORF CI-1 and ORF CI-2 are joined by removal of an intron in the region of overlap

A:Note: the exact position of the intron is given in reference S71838

R:Schalk, H.J.; Matzeit, V.; Schiller, B.; Schell, J.; Gronenborn, B.

EMBO J. 8, 359-364, 1989

A:Title: Wheat dwarf virus, a geminivirus of graminaceous plants needs splicing for replication

A:Reference number: S71838; MUID:89251559

A:Contents: annotation; intron position

C:Genetics:

A:Gene: CI

A:Introns: 210/3

C:Superfamily: tomato golden mosaic virus AL1 protein

C:Keywords: DNA replication

Query Match 83.3%; Score 30; DB 2; Length 351;

Best Local Similarity 57.1%; Pred. No. 86;

Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7

II I I

Db 266 KFTPNWK 272

RESULT 9

B24356

replication-associated protein - wheat dwarf virus (Swedish isolate)

N:Alternate names: ORF 30156/ORF 17292 composite protein

C:Species: wheat dwarf virus

A:Variety: Swedish isolate

C:Date: 09-Sep-1987 #sequence_revision 06-Dec-1996 #text_change 21-Aug-1998

C:Accession: B24356

R:Macdonald, S.W.; Macdonald, H.; Hamilton, W.D.O.; Courts, R.H.A.; Buck, K.W.

EMBO J. 4, 2173-2180, 1985

A:Title: The nucleotide sequence of cloned wheat dwarf virus DNA.

A:Reference number: A91012

A:Accession: B24356

A:Molecule type: DNA

A:Residues: 1-351 <MAC>

A:Cross-references: GB:X02869

A:Experimental source: Swedish isolate

A:Note: ORF 30156 and ORF 17292 are joined by removal of an intron in the region of overlap

A:Note: the exact position of the intron is given in reference S71838

R:Schalk, H.J.; Matzeit, V.; Schiller, B.; Schell, J.; Gronenborn, B.

EMBO J. 8, 359-364, 1989

A:Title: Wheat dwarf virus, a geminivirus of graminaceous plants needs splicing for replication

A:Reference number: S71838; MUID:89251559

A:Contents: annotation; intron position

C:Genetics:

A:Introns: 210/3

C:Superfamily: tomato golden mosaic virus AL1 protein

C:Keywords: DNA replication

Query Match 83.3%; Score 30; DB 2; Length 351;

Best Local Similarity 57.1%; Pred. No. 86;

Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7

II I I

Db 266 KFTPNWK 272

RESULT 10

GB3317

conserved hypothetical protein PA2630 [Imported] - Pseudomonas aeruginosa (strain PAO)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: GB3317

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; L.

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen

A:Reference number: AB2950; MUID:20437337

A:Accession: GB3317

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-389 <STO>

A:Cross-references: GB:AE004691; GB:AE004091; MID:g9948688; PIDN:AAC06018.1; GSPDB:GN

A:Experimental source: strain PAO1

C:Genetics:

A:Gene: PA2630

Query Match 83.3%; Score 30; DB 2; Length 389;

Best Local Similarity 57.1%; Pred. No. 95;

Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7

II I I

Db 110 RFLPSWR 116

RESULT 11

S38103

SP014 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YKR031c

C:Species: Saccharomyces cerevisiae

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 21-Jul-2000

C:Accession: S38103; S57804

R:Rustad, L.A.; Jauniaux, J.C.

submitted to the Protein Sequence Database, March 1994

A:Reference number: S38097

A:Molecule type: DNA

A:Residues: 1-1683 <UPR>

A:Cross-references: EMBL:228256; MID:g486462; PID:g486463; MIPS:YKR031c

A:Experimental source: strain S288C
R:Honigberg, S.M.; Conicella, C.; Esposito, R.E.
Genetics 130, 703-716, 1992
A:Title: Commitment to meiosis in *Saccharomyces cerevisiae*: involvement of the SP014 gene
A:Reference number: S57804; MUID:92258685
A:Accession: S57804
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 304-1683 <ION>
A:Cross-references: EMBL:L46807; NID:q951462; PIDN:AAA74938.1; PID:q954831
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
C:Genetics: S57804
A:Gene: SCD:SP014
A:Cross-references: SCD:S0001739; MIPS:YKK031c
A:Map position: 11R
C:Function:
A:Description: required for meiosis and spore formation

Query Match 83.3%; Score 30; DB 2; Length 1683;
Best Local Similarity 57.1%; Pred. No. 3; 7e-02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPXWR 7
II I I
DB 1227 KSPQWK 1233

RESULT 12
T39205
dynein light chain - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 19-May-2000
C:Accession: T39205
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Ransperger, U.; Pohl, T.
submitted to the EMBL Data Library, September 1999
A:Reference number: 221835
A:Accession: T39205
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-85 <MOO>
A:Cross-references: EMBL:AL110459; PIDN:CAB54155.1; GSPDB:GN000066; SPDB:SPAC926.07c
A:Experimental source: strain 972h; cosmid c926
C:Genetics:
A:Gene: SPDB:SPAC926.07c
A:Map position: 1
A:Introns: 9/3; 21/1; 32/3; 65/1
C:Superfamily: dynein light chain

Query Match 80.6%; Score 20; DB 2; Length 85;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPXW 6
II I I
DB 45 KSPWTW 50

RESULT 13
GB1198
hypothetical protein MM0438 [imported] - *Neisseria meningitidis* (strain MC58 serogroup
C:Species: *Neisseria meningitidis*
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: GB1198
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, P.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandli, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; V
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: GB1198

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <TET>
A:Cross-references: GB:AE002098; NID:g7225659; PIDN:AAF40876.1; PID:q722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0438

Query Match 80.6%; Score 29; DB 2; Length 122;
Best Local Similarity 57.1%; Pred. No. 50;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 KFXPXWR 7
II I I
DB 56 KFAPWQ 62

RESULT 14
CB1775
hypothetical protein NMA2047 [imported] - *Neisseria meningitidis* (strain 22491 serogr
C:Species: *Neisseria meningitidis*
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: CB1775
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
Hollard, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491
A:Reference number: AB1775; MUID:20222556
A:Accession: CB1775
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <PAR>
A:Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB05265.1; PID:g738
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA2047

Query Match 80.6%; Score 29; DB 2; Length 122;
Best Local Similarity 57.1%; Pred. No. 50;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPXWR 7
II I I
DB 56 KFAPWQ 62

RESULT 15
JQ1359
C2 protein - *Miscanthus streak virus*
C:Species: *Miscanthus streak virus*
C:Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 17-Nov-2000
C:Accession: JQ1359; JQ0922
R:Chatani, M.; Matsumoto, Y.; Mizuta, H.; Ikegami, M.; Boulton, M.I.; Davies, J.W.
J. Gen. Virol. 72, 2325-2331, 1991
A:Title: The nucleotide sequence and genome structure of the geminivirus *Miscanthus s*
A:Reference number: JQ1355; MUID:92013947
A:Accession: JQ1359
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <CHA>
A:Cross-references: DDBJ:D01030
A:Note: this reading frame extends between two stop codons and does not begin with a
R:Chatani, M.; Matsumoto, Y.; Mizuta, H.; Ikegami, M.; Boulton, M.I.; Davies, J.W.
submitted to JFIPID, May 1991
A:Description: The nucleotide sequence and genome structure of geminivirus *Miscanthus*
A:Reference number: JQ0918
A:Accession: JQ0922
A:Molecule type: DNA
A:Residues: 'M', 2-136 <CH2>
C:Comment: *Miscanthus streak virus* causes leaf streak.
C:Superfamily: tomato golden mosaic virus AL1 protein

Query Match 80.6%; Score 29; DB 2; Length 136;
Best Local Similarity 57.1%; Pred. No. 56;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPXWR 7
1111:
Db 59 KFLPQWK 65

Search completed: February 27, 2002, 11:52:59
Job time: 451 sec

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Poliovirus.
 OX NCBI_TaxID-12048;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89171329; PubMed=2466700;
 RA van der Wiik F., Huisman M.J., Cornelissen R.J.C., Huttinga H.,
 RA Goldbach R.W.;
 RT "Nucleotide sequence and organization of potato leafroll virus
 RT genomic RNA";
 RL FEBS Lett. 245:51-56(1989).
 CC -!- SIMILARITY: 54% SIMILARITY TO BEET WESTERN YELLOWS VIRUS ORF2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y07496; CAA68795.1;
 DR PIR: S03547; S03547.
 DR InterPro: IPR000382; Luteo_ORF2.
 DR Pfam: PF02122; Luteo_ORF2; 1.
 DR PRINTS: PR00913; LVIRUSORF2.
 SQ SEQUENCE 639 AA: 69676 MW: 92E1473FE3FEF148 CRC64;

Query Match 91.78; Score 33; DB 1; Length 639;
 Best Local Similarity 71.4%; Pred. No. 21;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPXR 7
 |||
 DB 614 KFIQSWR 620

RESULT 3
 RBL2_RHOSH
 ID RBL2_RHOSH STANDARD; PRT; 459 AA.
 AC P29278;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE RIBULOSE BISPHOSPHATE CARBOXYLASE LARGE CHAIN 2 (EC 4.1.1.39)
 DE (RUBISCO LARGE SUBUNIT).
 DE CBBL2 OR RBPL.
 CN RBL2_RHOSH
 OS Rhodospirillum rubrum (Rhodospirillum rubrum sphaeroides).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
 OC Rhodospirillum.
 OX NCBI_TaxID-1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wagner S.J., Stevens S.E. Jr., Nixon B.T., Lambert D.H.,
 RA Quivey R.G. Jr., Tabita F.R.;
 RT "Nucleotide and deduced amino acid sequence of the Rhodospirillum
 RT sphaeroides gene encoding form II ribulose-1,5-bisphosphate
 RT carboxylase/oxygenase and comparison with other deduced forms I and II
 RT sequences";
 RT FEBS Microbiol. Lett. 55:217-222(1988).
 RL [2]
 RN SEQUENCE OF 1-36 FROM N.A.
 RP MEDLINE=92041881; PubMed=1939098;
 RX Chen J.-H., Gibson J.L., McCue L.A., Tabita F.R.;
 RT "Identification, expression, and deduced primary structure of
 RT transketolase and other enzymes encoded within the form II CO2
 RT fixation operon of Rhodospirillum rubrum sphaeroides";
 RL J. Biol. Chem. 266:20447-20452(1991).
 CC -!- FUNCTION: RUBISCO CATALYZES TWO REACTIONS: THE CARBOXYLATION OF
 CC D-RIBULOSE 1,5-BISPHOSPHATE, THE PRIMARY EVENT IN PHOTOSYNTHETIC
 CC CARBON DIOXIDE FIXATION, AS WELL AS THE OXIDATIVE FRAGMENTATION OF
 CC THE PENTOSE SUBSTRATE IN THE PHOTORESPIRATION PROCESS. BOTH

CC REACTIONS OCCUR SIMULTANEOUSLY AND IN COMPETITION AT THE SAME
 CC ACTIVE SITE.
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + CO(2) -
 CC 2 3-PHOSPHO-D-GLYCERATE.
 CC -!- CATALYTIC ACTIVITY: D-RIBULOSE 1,5-BISPHOSPHATE + O(2) -
 CC 3-PHOSPHO-D-GLYCERATE + 2-PHOSPHOGLYCOLATE.
 CC -!- SUBUNIT: IN CONTRAST TO FORM I GENES, THE FORM II RUBISCO ARE
 CC COMPOSED OF ONLY LARGE SUBUNITS.
 CC -!- MISCELLANEOUS: THIS PROTEIN IS ENCODED WITHIN THE FORM II
 CC RIBULOSE-BISPHOSPHATE CARBOXYLASE OPERON.
 CC -!- SIMILARITY: BELONGS TO THE RUBISCO LARGE CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

EMBL: M68914; AAA26158.1; ALT_SEQ.
 DR PIR: E41080; E41080.
 DR HSPP: P04718; IRBA.
 DR InterPro: IPR000685; RUBISCO_large.
 DR Pfam: PF00016; RUBISCO_large; 1.
 DR PROSITE: PS00157; RUBISCO_LARGE; 1.
 KW Phytosynthesis; Carbon dioxide fixation; Photorespiration;
 KW Lyase; Oxidoreductase; Monooxygenase; Multigene family.
 FT ACT_SITE 191
 FT BINDING OF CO(2) ACTIVATES THE ENZYME.
 SQ SEQUENCE 459 AA: 50519 MW: 299ABAA836B0683E CRC64;

Query Match 88.98; Score 32; DB 1; Length 459;
 Best Local Similarity 71.4%; Pred. No. 24;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPXR 7
 |||
 DB 446 KFIQSWR 452

RESULT 4
 IF37_ARATH
 ID IF37_ARATH STANDARD; PRT; 591 AA.
 AC P55820;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PUTATIVE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT 7 (EIF-3
 DE ZEFA).
 DE AT4G20980 OR T13K14.140.
 CN Arabidopsis thaliana (Mouse-ear cress).
 OS Arabidopsis thaliana (Brassicaceae).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083488; PubMed=10617198;
 RA Hayer K.F.X., Schueller C., Wambutt K., Murphy G., Voelckert G.,
 RA Hill T., Duesterhoeft A., Stiekema W., Kntian K.-D., Terry N.,
 RA Harris B., Ansgore W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis M., Delany M., Pulgomech P., Watson M., Schmidheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Hilham L., Robben J.,
 RA Van der Schuren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Bracken M., Welljens I., Voet M., Bastiaens I., Aert R., DeLoor E.,
 RA Weizenecker T., Bolhe G., Ramsperger U., Hilbert H., Braun M.,
 RA Heger E., Brandt A., Peters S., van Staveren M., Dirkse W.,
 RA Noolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,

CN SP014 OR PL11 OR YKR031C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN [1]
 RN [2]
 RN [3]
 RN [4]
 RN [5]
 RN [6]
 RN [7]
 RN [8]
 RN [9]
 RN [10]
 RN [11]
 RN [12]
 RN [13]
 RN [14]
 RN [15]
 RN [16]
 RN [17]
 RN [18]
 RN [19]
 RN [20]
 RN [21]
 RN [22]
 RN [23]
 RN [24]
 RN [25]
 RN [26]
 RN [27]
 RN [28]
 RN [29]
 RN [30]
 RN [31]
 RN [32]
 RN [33]
 RN [34]
 RN [35]
 RN [36]
 RN [37]
 RN [38]
 RN [39]
 RN [40]
 RN [41]
 RN [42]
 RN [43]
 RN [44]
 RN [45]
 RN [46]
 RN [47]
 RN [48]
 RN [49]
 RN [50]
 RN [51]
 RN [52]
 RN [53]
 RN [54]
 RN [55]
 RN [56]
 RN [57]
 RN [58]
 RN [59]
 RN [60]
 RN [61]
 RN [62]
 RN [63]
 RN [64]
 RN [65]
 RN [66]
 RN [67]
 RN [68]
 RN [69]
 RN [70]
 RN [71]
 RN [72]
 RN [73]
 RN [74]
 RN [75]
 RN [76]
 RN [77]
 RN [78]
 RN [79]
 RN [80]
 RN [81]
 RN [82]
 RN [83]
 RN [84]
 RN [85]
 RN [86]
 RN [87]
 RN [88]
 RN [89]
 RN [90]
 RN [91]
 RN [92]
 RN [93]
 RN [94]
 RN [95]
 RN [96]
 RN [97]
 RN [98]
 RN [99]
 RN [100]

DR	36001732; JPOA; 1	
DR	InterPro: IPR001849; PH.	
DR	InterPro: IPR001736; PLD.	
DR	InterPro: IPR001683; PX.	
DR	pfam: PF00614; PLDC; 2.	
DR	SMART; SM00233; PH; 1.	
DR	SMART; SM00155; PLDC; 2.	
DR	SMART; SM00312; PX; 1.	
KW	Hydrolase; Lipid degradation; Sporulation; Meiosis.	
Fr	DMA; N 102 107	POLY-ASN.
Fr	C-CHAIN 325 330	POLY-LEU.
SO	SEQUENCE 1380 AA: 160994 MW: 594C48DCAT07A3DB CRC64;	

```

Query Match      83.3%  Score 30;  DB 1;  Length 1380;
Best Local Similarity 57.1%  Pred. No. 1.6e+02;
Matches: 4;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0;

Cq      1  KFXPXR 7
      ||| |

```

```

DB 924 KFSHOWK 930

RESULT 6
DYLLI_SCHPO STANDARD; PRT: 85 AA.
AC O9UR05:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE DYNEIN LIGHT CHAIN 1, CYTOPLASMIC.
GN DL2 OK SPAC926.07C.
OS Schizosaccharomyces pombe (fission yeast).
CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC NCBI_TaxID=4896;
CC SEQUENCE FROM N.A.
KP STRAIN=972;
RA Miki F., Shimanuki M., Okazaki K., Niwa O.;
RT "Role of TCYEX-1 like dynein light chain for the function of telomere
KL associated-spindle pole body in fission yeast.";
KL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
KN [2]
KP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Ransperger U., Pohl T.;
KL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases
CC -1- FUNCTION: MAY BE INVOLVED IN SOME ASPECTS OF DYNEIN-RELATED
CC INTRACELLULAR TRANSPORT AND MOTILITY. MAY PLAY A ROLE IN CHANGING
CC OR MAINTAINING THE SPATIAL DISTRIBUTION OF CYTOSKELETAL
CC STRUCTURES (BY SIMILARITY).
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN LIGHT CHAIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF197476; AAF05842.1; -
CC ENHL: A110469; CAB54155.1; -
CC HSP: Q15701; I8K0
CC InterPro: IPR001372; Dynein_light.
CC Pfam: PF01221; Dynein_light; 1.
CC ProDom: PD005145; Dynein_light; 1.
CC ProSITE: PS01239; DYNEIN_LIGHT_1; 1.
CC Motor protein; Microtubules; Dynein.
KW SEQUENCE 85 AA; 9822 MW; 81B5EC20081D628A CRC64;
SQ

```

```

Query Match 80.6% Score 29; DB 1: Length 85;
Best Local Similarity 66.7% Pred. No. 20;
Matches 4: Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXIXW 6
II I I
DB 45 KESPTW 50

RESULT 7
Y17K_TYDVA STANDARD; PRT: 146 AA.
AC P31618;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE DYNEIN LIGHT CHAIN 1, CYTOPLASMIC.
GN DL2 OK SPAC926.07C.
OS Schizosaccharomyces pombe (fission yeast).
CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC NCBI_TaxID=4896;
CC SEQUENCE FROM N.A.
KP STRAIN=972;
RA Miki F., Shimanuki M., Okazaki K., Niwa O.;
RT "Role of TCYEX-1 like dynein light chain for the function of telomere
KL associated-spindle pole body in fission yeast.";
KL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
KN [2]
KP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Ransperger U., Pohl T.;
KL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases
CC -1- FUNCTION: MAY BE INVOLVED IN SOME ASPECTS OF DYNEIN-RELATED
CC INTRACELLULAR TRANSPORT AND MOTILITY. MAY PLAY A ROLE IN CHANGING
CC OR MAINTAINING THE SPATIAL DISTRIBUTION OF CYTOSKELETAL
CC STRUCTURES (BY SIMILARITY).
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN LIGHT CHAIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF197476; AAF05842.1; -
CC ENHL: A110469; CAB54155.1; -
CC HSP: Q15701; I8K0
CC InterPro: IPR001372; Dynein_light.
CC Pfam: PF01221; Dynein_light; 1.
CC ProDom: PD005145; Dynein_light; 1.
CC ProSITE: PS01239; DYNEIN_LIGHT_1; 1.
CC Motor protein; Microtubules; Dynein.
KW SEQUENCE 85 AA; 9822 MW; 81B5EC20081D628A CRC64;
SQ

```

```

DE HYPOPHETICAL 16.7 KDA PROTEIN.
GN C2.
OS Tobacco yellow dwarf virus (strain Australia) (TYDV).
OX Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=31599;
KN [1]
KP SEQUENCE FROM N.A.
XX MEDLINE=92188538; PubMed=1546458;
RA Morris B.A.M., Richardson K.A., Halsey A., Zhan X., Thomas J.E.;
RT "The nucleotide sequence of the infectious cloned DNA component of
KL tobacco yellow dwarf virus reveals features of geminiviruses
KL infecting monocotyledonous plants.";
KL Virology 187:633-642(1992).
CC -1- FUNCTION: MAY BE INVOLVED IN REPLICATION.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M81103; AAA47949.1; ALT_INIT.
CC FIC: C42452; C42452.
CC InterPro: IPR001191; Gemini_AL1.
CC Pfam: PF00799; Gemini_AL1; 1.
CC ProDom: PD000736; Gemini_AL1; 1.
CC DNA replication.
KW SEQUENCE 146 AA; 16743 MW; 9C8B3C87508D010A CRC64;
SQ

```

```

Query Match 80.6% Score 29; DB 1: Length 146;
Best Local Similarity 57.1% Pred. No. 33;
Matches 4: Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXIXW 7
II I I
DB 68 KFXPLWK 74

RESULT 8
YHET_ECOLI STANDARD; PRT: 340 AA.
AC F4524;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE HYPOPHETICAL PROTEIN YHET.
GN YHET OR H3353.
OS Escherichia coli.
CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_TaxID=562;
KN [1]
KP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
XX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
XX "The complete genome sequence of Escherichia coli K-12.";
XX Science 277:1232-1244(1997).
CC -1- SIMILARITY: BELONGS TO THE UPF0017 FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC

```



```
CC EMBL: U18997; AAC58150.1; ..
DR EMBL: AE000411; AAC76378.1; ..
DR Ecogen: EG12904; yhet.
DR InterPro: IPR000073; Abhydrolase.
DR InterPro: IPR000379; Est_lip_thioest_actsite.
DR InterPro: IPR000952; UPF0017.
DR Pfam: PF00561; abhydrolase; 1.
DR PROSITE: PS01133; UPF0017; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 147 164
SQ SEQUENCE 340 AA; 38495 MW; 7DB6BA6795CBA64D CRC64;

Query Match      80.6%; Score 29; DB 1; Length 340;
Best Local Similarity 57.1%; Pred. No. 70;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPWXR 7
DB 42 KFTPYWQ 48

RESULT 9
ID CYB_DROME: STANDARD; PRT; 378 AA.
AC P18935;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOCHROME B;
GN MT:CYT-B OR COX OR CYTB.
OS Drosophila melanogaster (Fruit fly).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Phylorhoda; Prosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRETAGNE;
RX MEDLINE=88212147; PubMed=3130291;
RA Garesse R.;
RT "Drosophila melanogaster mitochondrial DNA: gene organization and
Genetics 118:649-663(1988).";
RL CYTOCHROME B;
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS.
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN.
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M37275; AAC69714.1; ..
DR EMBL: U37541; AAC47822.1; ..
DR PIR: S01190; S01190.
DR FlyBase: FBgn0013678; mt:Cyt-b.
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c1.
DR Pfam: PF00033; cytochrome_b_c1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_H_HO; 1.

4W Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Hemo.
FT METAL 84 84 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 98 98 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 183 183 IRON 2 (HEME B562 AXIAL LIGAND).
FT METAL 197 197 IRON 1 (HEME B566 AXIAL LIGAND).
SQ SEQUENCE 378 AA; 43224 MW; 6CB9067643ABCE83 CRC64;

Query Match      80.6%; Score 29; DB 1; Length 378;
Best Local Similarity 66.7%; Pred. No. 77;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPW 6
DB 19 KFTPTW 114

RESULT 10
ID SPS2_MOUSE: STANDARD; PRT; 452 AA.
AC P97364;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE SELENIDE, WATER KINASE 2 (EC 2.7.9.3) (SELENOPHOSPHATE SYNTHETASE 2)
DE (SELENIUM DONOR PROTEIN 2).
GN SPS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017645; PubMed=7588067;
RA Guimaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,
RA Lee F., McClanahan T.;
RT "A new approach to the study of haematopoietic development in the
RT yolk sac and embryoid bodies.";
RL Development 121:3335-3346(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97140286; PubMed=8986768;
RA Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,
RA Zlotnik A.;
RT "Identification of a novel self homolog from eukaryotes, bacteria,
RT and archaea: is there an autoregulatory mechanism in selenocysteine
RT metabolism?";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).
CC -!- FUNCTION: SYNTHESIZES SELENOPHOSPHATE FROM SELENIDE AND ATP.
CC -!- CATALYTIC ACTIVITY: ATP + SELENIDE + H(2)O = AMP + SELENOPHOSPHATE
CC + PHOSPHATE.
CC -!- COFACTOR: SELENOCYSTEINE. THE ACTIVE-SITE SELENOCYSTEINE IS
CC ENCODED BY THE OPAL CODON, UGA.
CC -!- SIMILARITY: BELONGS TO THE SELENOPHOSPHATE SYNTHETASE 1 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U43285; AAC53024.1; ..
DR EMBL: MG1108388; SPS2.
DR InterPro: IPR000728; AIRS_related.
DR Pfam: PF00596; AIRS; 1.
DR KX: Transferase; Selenium; Selenocysteine; ATP-binding.
FT ACT_SITE 63 63
FT SE_CYS 63 63
FT POTENTIAL..
```

PT SITE 66 66 IMPORTANT FOR CATALYTIC ACTIVITY (BY
SIMILARITY).
FT NP_BIND 322 328 ATP (POTENTIAL).
FT DOMAIN 2 9 POLY-ALA.
FT DOMAIN 433 440 POLY-ALA.
SO SEQUENCE 452 AA; 47786 MW; 9DA6F7250CFE80E4 CRC64;

Query Match 80.6%; Score 29; DB 1; Length 452;
Best Local Similarity 66.7%; Pred. No. 91;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 FXPMWR 7.
I I I I
Db 48 FSPSWR 53

RESULT 11

YZ64_SYNY3 STANDARD; PRT; 477 AA.
AC P73436;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 54.0 KDA PROTEIN SLI1464.
GN SLI1464.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN 111
SEQUENCE FROM N.A.
RP MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kofani H., Tanaka A., Asamizu E., Nakamura Y.,
Miyaajima N., Hiroseawa M., Sugliura M., Sasamoto S., Kimura T.,
Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
Yamada M., Yasuda M., Tabata S.;
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE UPF0061 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: D90906; BAA17476.1;
DR InterPro: IPR001846; UPF0061.
IK Pfam: PF02696; UPF0061.1.
SQ SEQUENCE: 477 AA; 54041 MW; 81F6899H1A6D613C CRC64;

Query Match 80.6%; Score 29; DB 1; Length 477;
Best Local Similarity 66.7%; Pred. No. 95;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 FXPMWR 7
I I I I
Db 388 FSPSWR 393

RESULT 12

AAPI_YEAST STANDARD; PRT; 856 AA.
AC P37898;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ALANINE/ARGININE AMINOPEPTIDASE (EC 3.4.11.-).
GN AAP1 OR YHR047C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN 111
SEQUENCE OF 107-652 FROM N.A.
RP MEDLINE=93300827; PubMed=8100228;
RA Caprioglio D.R., Padilla C., Werner-Washburne M.;
RT "Isolation and characterization of AAP1. A gene encoding an
alanine/arginine aminopeptidase in yeast.";
RL J. Biol. Chem. 268:14310-14315(1993).
RN 121
SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RC MEDLINE=94378003; PubMed=8091229;
RX Johnston M., Andrews S., Brinkman K., Cooper J., Ding H., Dover J.,
Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
Vignati D., Wilcox L., Wohlman P., Waterston K., Wilson R.,
Vandini M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
VIII.";
RL Science 265:2077-2082(1994).
CC -1- FUNCTION: POSITIVE EFFECTOR OF GLYCOGEN ACCUMULATION. MAY BE
INVOLVED IN NUTRIENT-SENSING.
CC -1- COFACTOR: BINDS ONE ZINC ION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M1 (ZINC METALLOPROTEASE);
ALSO KNOWN AS THE PERP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: L12542; NOT_ANNOTATED_CDS.
DR EMBL: U00062; AAB68919.1;
DR PIR: S46750; S46750.
DR MEGOPS: M01.007;
DR SGE: S0001089; AAP1.
DR InterPro: IPR001930; Aladiptase.
DR InterPro: IPR000130; Zn_Mtpeptdse.
DR Pfam: PF01433; Peptidase_M1; 1.
DR PRINTS: PR00756; ALADIPTASE.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc.
FT METAL 300 300 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 301 301 BY SIMILARITY.
FT METAL 304 304 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 323 323 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 386 386 PROTON DONOR (POTENTIAL).
FT CONFLICT 550 569 MISSING (IN REF. 2).
FT CONFLICT 646 646 V -> E (IN REF. 2).
SQ SEQUENCE 856 AA; 97662 MW; 0209F9E0298DF9DB CRC64;

Query Match 80.6%; Score 29; DB 1; Length 856;
Best Local Similarity 57.1%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPMWR 7
I I I I
Db 336 KFXPMWR 342

RESULT 13

```

Query Match      77.8%; Score 28; DB 1; Length 151;
Best Local Similarity 66.7%; Pred. No. 53;
Mat. 100%; 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CY      2 EXPXWR 7
      1 1 1 1
DB      6 ESWLR 11

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use.
CC -----
CC

Result No.	Score	Query Match	LenqLh	DB	ID	Description
1	33	91.7	241	5	Q9V1D3	Q9V1D3 drosophila
2	33	91.7	250	2	Q9FUM2	Q9FUM2 corynebacte
3	33	91.7	639	12	Q84836	Q84836 potato leaf
4	32	88.9	348	2	Q9F756	Q9F756 bacteroides
5	32	88.9	380	1	Q57935	Q57935 methanococ
6	32	88.9	462	10	Q23080	Q23080 arabidopsis
7	32	88.9	588	10	Q9FKV6	Q9FKV6 arabidopsis
8	32	88.9	688	12	Q99H27	Q99H27 helicoverp
9	31	86.1	254	1	Q9YBB9	Q9YBB9 aetopyrum p
10	30	83.3	150	12	Q67622	Q67622 wheat dwarf
11	30	83.3	160	10	Q9M7W6	Q9M7W6 arabidopsis
12	30	83.3	254	2	Q9F1E6	Q9F1E6 acetobacter
13	30	83.3	303	10	Q9LS14	Q9LS14 arabidopsis
14	30	83.3	317	10	Q80627	Q80627 arabidopsis
15	30	83.3	389	2	Q910R8	Q910R8 pseudomonas
16	30	83.3	407	2	Q9AA59	Q9AA59 caulobacter
17	30	83.3	490	2	Q9A3L3	Q9A3L3 caulobacter
18	30	83.3	493	5	Q9N8B9	Q9N8B9 plasmodium
19	24	80.6	73	12	Q9DSV9	Q9DSV9 ascovirus d

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.K., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Stimpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas K., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 HL Science 287:2185-2195(2000).
 DR EMBL: AE003624; AAF52759.1;
 DR FlyBase: FBgn0033107; CG3694.
 SQ SEQUENCE 241 AA; 27071 MW; 200925FE35C3F733 CRC64;

Query Match 91.7%; Score 33; DB 5; Length 241;
 Best Local Similarity 71.4%; Pred. No. 46;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7
 ||| ||
 DB 67 KFTPVWR 73

RESULT 2
 Q9EUM2 PRELIMINARY; PRT; 250 AA.
 ID Q9EUM2;
 AC Q9EUM2;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE YC4L.
 GN YC4L.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Plasmid R-plasmid pCG4.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 CC Corynebacterium.
 OX NCHI_TaxID=1718;
 RN |||
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC31830;
 RA Tauch A., Puschler A., Kalinowski J.;
 RT "DNA sequence and genetic organization of the integron-carrying R-
 plasmid pCG4 of *Corynebacterium glutamicum*.";
 RL Submitted (Jul-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF164956; AAC00294.1;
 DR InterPro: IPR002145; CopG_UtlH_4.
 DR Pfam: PF01402; UtlH_4; 1.
 KW Plasmid.
 SQ SEQUENCE 250 AA; 28804 MW; 195C86367H4B3157 CRC64;

Query Match 91.7%; Score 33; DB 2; Length 250;
 Best Local Similarity 71.4%; Pred. No. 47;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7
 ||| ||
 DB 213 KFKPAWR 219

RESULT 3
 Q84836 PRELIMINARY; PRT; 639 AA.
 ID Q84836;
 AC Q84836;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE GRE 2.
 OS Potato leafroll virus (PLRV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
 OC Poterovirus.
 OX NCHI_TaxID=12045;
 RN |||
 RP SEQUENCE FROM N.A.
 RC STRAIN=THE POLISH ISOLATE;
 RX MEDLINE=95250405; PubMed=7732757;
 RA Palucha A., Sadowy E., Kujawa A., Juszczyk M., Zagorski W.,
 RA Hulanicka D.;
 RT "Nucleotide sequence of RNA of a Polish isolate of potato leafroll
 luteovirus.";
 RL Acta Biochim. Pol. 41:405-414(1994).
 DR EMBL: X74789; CAA52790.1;
 DR InterPro: IPR000382; Luteo_ORF2.
 DR Pfam: PF02122; Luteo_ORF2; 1.
 DR PRINTS: PR00913; LVIRUSORF2.
 SQ SEQUENCE 639 AA; 69649 MW; 6EBA3EA9C55016B3 CRC64;

Query Match 91.7%; Score 33; DB 12; Length 639;
 Best Local Similarity 71.4%; Pred. No. 11e+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7
 ||| ||
 DB 514 KFP5WR 620

RESULT 4
 Q9F756 PRELIMINARY; PRT; 348 AA.
 ID Q9F756;
 AC Q9F756;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE PUTATIVE GLYCOSYLTRANSFERASE.
 GN WCQ.
 OS Bacteroides fragilis.
 CC Bacteria; CF group; Bacteroidaceae; Bacteroides.
 OX NCHI_TaxID=817;
 RN |||
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 9343;
 RX MEDLINE=20490546; PubMed=11035722;
 RA Coyne M.J., Kalka-Moll W., Tzianabos A.O., Kasper D.L., Comstock L.E.;
 RT "Bacteroides fragilis NCTC9343 produces at least three distinct
 Capsular Polysaccharides: Cloning, Characterization, and Reassignment
 of Polysaccharide B and C Biosynthesis Loci.";
 RL Infect. Immun. 68:6176-6181(2000).
 DR EMBL: AF285774; AAC26474.1;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Transferase.
 SQ SEQUENCE 348 AA; 40437 MW; A259548CC4C8AA5B CRC64;

Query Match 88.9%; Score 32; DB 2; Length 348;
 Best Local Similarity 71.4%; Pred. No. 1e+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXWR 7
 ||| ||
 DB 299 KFLPYWR 305

RESULT 5
 Q57935 PRELIMINARY; PRT; 380 AA.
 ID Q57935;
 AC Q57935;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE FORMATE HYDROGENLYASE: SUBUNIT 5 (PFI, SUBUNIT 5) (HYDROGENASE-
 GN 3 COMPONENT E).
 DE MJ0515.
 OS Methanococcus jannaschii.
 OS Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
 OC Methanococcus.
 OX NCBI_Taxid=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Cocayne J.D.,
 RA Kervilange A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -I- COFACTOR: NICKEL (BY SIMILARITY).
 CC -I- PATHWAY: HYDROGEN METABOLISM; FHL PATHWAY (BY SIMILARITY).
 CC -I- SUBUNIT: FHL COMPRISES OF A FORMATE DEHYDROGENASE, UNIDENTIFIED
 CC ELECTRON CARRIERS AND A HYDROGENASE (ISOZYME 3). IN THIS NON-
 CC ENERGY CONSERVING PATHWAY MOLECULAR HYDROGEN AND CARBODIOXIDE FROM
 CC FORMATE ARE RELEASED (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE COMPLEX I 49 KDA SUBUNIT FAMILY.
 DR EMBL: U67501; AA898504.1;
 DR TIGR: MJ0515;
 DR InterPro: IPR001135; Complex1_49kd.
 DR InterPro: IPR001501; NiFeSe_Hases.
 DR Pfam: PF00346; complex1_49kd; 1.
 DR Pfam: PF00374; NiFeSe_Hases; 1.
 DR PROSITE: PS00535; COMPLEX1_49K; UNKNOWN_1.
 DR PROSITE: PS00507; NI_HGRNASH_1; UNKNOWN_1.
 DR Hypothetical protein: Oxidoreductase; NAD: Iron-sulfur; 4Fe-4S;
 KW Nickel; Complete proteome.
 SQ SEQUENCE 380 AA; 43638 MW; 3EAD16D67834A76F CRC64;

Query Match 88.9%; Score 32; DB 1; Length 380;
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPKWR 7
 DB 240 KFKPVWR 246

RESULT 6
 ID Q23080 PRELIMINARY; PRT: 462 AA.
 AC Q23080;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE SIMILARITY TO GAG PROTEINS (AT4G00980 PROTEIN).
 GN A_TM018A10.1 OR AT4G00980.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OS Eukaryota; Viridiplantae; Streptophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, COLUMBIA;
 RA Dempsey S., Harper M.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN-CV, COLUMBIA;
 RA Waterston R.;
 PL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lamar B., Stoneking T., Stumpf J., Mewes H.W., Lemcke K.,
 RA Harper K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA CH Arabidopsis sequencing project;
 RC Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -I- SIMILARITY: BELONGS TO 2N-FINGER CCHC TYPE FAMILY.
 DR EMBL: AF013294; AA862847.1;
 DR EMBL: AL161491; CAB80907.1;
 DR InterPro: IPR000531; TonB_boxC.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam: PF00098; zf-CCHC; 1.
 DR PRINTS: PR00939; C2HCZNFINGER.
 DR SMART: SM00343; Znf_C2HC; 1.
 DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
 KW zinc-finger.
 SQ SEQUENCE 462 AA; 52052 MW; C4A12045D0376E33 CRC64;

Query Match 88.9%; Score 32; DB 10; Length 462;
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPKWR 7
 DB 127 KFKPSWR 333

RESULT 7
 ID Q9PKV6 PRELIMINARY; PRT: 588 AA.
 AC Q9PKV6;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE EUKARYOTIC TRANSLATION INITIATION FACTOR 3 SUBUNIT 7.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Soemmatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=98344145; PubMed=9679202;
 RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asanizu E., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
 RT features of the regions of 1,381,565 bp covered by twenty one
 RT physically assigned B1 and TAC clones.";
 RL DNA Res. 5:131-145(1998).
 DR EMBL: AB011475; BAB10117.1;
 KW Initiation factor.
 SQ SEQUENCE 588 AA; 66223 MW; 2294E8AF3F6A942H CRC64;

Query Match 88.9%; Score 32; DB 10; Length 588;
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPKWR 7
 DB 118 KFKPKWR 124

RESULT 8
 Q99H2

ID O99H27 PRELIMINARY; PRT: 688 AA.
 AC O99H27;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE P74.
 OS Helicoverpa armigera nucleopolydnavirus G4.
 CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 CC Nucleopolydnavirus.
 ON NCBI_TaxID=148363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Deng F., Chen X., Viak J.M., Arif B.M., Hu Z.;
 RA "Sequence analysis of the gp37 gene of Heliothis armigera single-
 RT nucleocapsid nucleopolydnavirus.";
 RT Zhongguo Bingdaxue 15:35-42(2000).
 RL [2]
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Wang H., Hu Z., Sun X., Viak J.M., Chen X.;
 RA "Sequence analysis of the Iap3 gene of Heliothis armigera single-
 RT nucleocapsid nucleopolydnavirus.";
 RT Zhongguo Bingdaxue 15:43-49(2000).
 RL [3]
 RN [4]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21078302; PubMed=11210934;
 RA Wang H., Chen X., Wang H., Arif B.M., Viak J.M., Hu Z.;
 RA "Nucleotide sequence and transcriptional analysis of a putative basic
 RT DNA-binding protein of Helicoverpa armigera polydnavirus.";
 RT Virus Genes 22:113-120(2001).
 RL [4]
 RN [5]
 RP SEQUENCE FROM N.A.
 RA PubMed=11125177;
 RA Chen X., Ijkel W.F., Tarchini R., Sun X., Sandbrink H., Wang H.,
 RA Peters S., Zuidema D., Lankhorst R.K., Viak J.M., Hu Z.;
 RA "The sequence of the Helicoverpa armigera single-nucleocapsid
 RT nucleopolydnavirus genome.";
 RT J. Gen. Virol. 82:241-257(2001).
 RL [5]
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Chen X., Ijkel W.F., Tarchini R., Sun X., Sandbrink H., Wang H.,
 RA Peters S., Zuidema D., Lankhorst R.K., Viak J.M., Hu Z.;
 RA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF271059; AAG53763.1;
 SQ SEQUENCE 688 AA; 78434 MW; BDE610E6D16AC6F5 CRC64;

Query Match 88.9%; Score 32; DB 12; Length 688;
 Best Local Similarity 71.4%; Pred. No. 1.9e+02;
 Matches 5; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPKWR 7
 DB 49 KFXPKWR 55

RESULT 9
 Q9YBB9 PRELIMINARY; PRT: 254 AA.
 ID Q9YBB9
 AC Q9YBB9;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE HYPOThETICAL. 27.0 KDa PROTEIN APR1678.
 CN APR1678.
 OS Aeropyrum pernix.
 CC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 CC Aeropyrum.
 ON NCBI_TaxID=56636;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=KI;
 RX MEDLINE=99310339; PubMed=10382966;
 RA Kawarabayashi Y., Iino Y., Horikawa H., Yamazaki S., Iwakawa Y.,

RA Jia-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
 RT "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000062; BAA80679.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 254 AA; 27026 MW; D173589517FD0D2A CRC64;

Query Match 86.1%; Score 31; DB 1; Length 254;
 Best Local Similarity 57.1%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPKWR 7
 DB 168 KFXPAWK 174

RESULT 10
 Q67622 PRELIMINARY; PRT: 150 AA.
 ID Q67622
 AC Q67622;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE ORF 1742.
 OS Wheat dwarf virus (WDV).
 CC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
 ON NCBI_TaxID=10834;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ma-Dowell S.W., MacDonald H., Hamilton W.D.O., Coutlis R.H.A.,
 RA Buck K.W.;
 RL EMBL J. 4:2173-2180(1985).
 DR EMBL: X02869; CAA26624.1;
 LR InterPro: IPR001191; Gemini_AL1.
 DR Pfam: PF00799; Gemini_AL1; 1.
 DR EC:Dom: PD000736; Gemini_AL1; 1.
 SQ SEQUENCE 150 AA; 17292 MW; 3B7394DCB793F9FB CRC64;

Query Match 83.3%; Score 30; DB 12; Length 150;
 Best Local Similarity 57.1%; Pred. No. 1.1e+02;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPKWR 7
 DB 65 KFXPKWR 71

RESULT 11
 Q9M7W5 PRELIMINARY; PRT: 160 AA.
 ID Q9M7W5
 AC Q9M7W5;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE M4J6.1 PROTEIN.
 CN M4J6.1.
 CS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Denito M., Creasy T.H., Haas B., Wu D.,
 RA Maiti R., Ronning C.M., Kuo H., Fujii C.Y., Ulterback T.R.,

RA Harstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome III p1 MCH6 genomic sequence."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC024128; AAF35940.1;
 SQ SEQUENCE 160 AA; 18620 MW; 95651977DE161826 CRC64;

Query Match 83.3%; Score 30; DB 10; Length 160;
 Best Local Similarity 57.1%; Pred. No. 1.2e+02;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KXPXWR 7
 ID 48 KYTPWR 54

RESULT 12
 O9F1E6 PRELIMINARY; PRT; 254 AA.

AC O9F1E6;
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MODC.
 GN MODC.
 OS Acetobacter diazotrophicus.
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 CC Gluconacetobacter.
 OX NCBI_TaxID=33996;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAL5;
 RA Sevilla M.O., de Oliveira A., Baldani I., Kennedy C.;

RT "Contributions of the bacterial endophyte Acetobacter diazotrophicus
 RT to sugarcane nutrition: A preliminary study."
 RL Symbiosis (Philadelphia) 25:181-192(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=PAL5;
 RA Teixeira K.R.S., Wuelling M., Morgan T., Galler R., Zellerman E.,
 RA Baldani J.I., Kennedy C., Meletzus D.;
 RT "Molecular analysis of the chromosomal region encoding the nifA and
 RT nifB genes of Acetobacter diazotrophicus."
 RL FEMS Microbiol. Lett. 176:301-309(1999).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN=PAL5;
 RX MEDLINE=20545459; PubMed=11092875;
 RA Lee S., Keth A., Meletzus D., Sevilla M., Kennedy C.;
 RT "Characterization of a Major Cluster of nif, fix, and Associated Genes
 RT in a Sugarcane Endophyte, Acetobacter diazotrophicus."
 RL J. Bacteriol. 182:7088-7091(2000).
 DR EMBL: AF030414; AAG35385.1;
 DR InterPro: IPR003593; AAA.
 DR InterPro: IPR003439; ABC-transporter.
 DR InterPro: IPR001687; ATP-GTP_A.
 DR Pfam: PF00005; ABC_tran; 1.
 DR SMART: SM00382; AAA; 1.
 SQ SEQUENCE 254 AA; 27729 MW; BD96E72DBBC2BC85 CRC64;

Query Match 83.3%; Score 30; DB 2; Length 254;
 Best Local Similarity 57.1%; Pred. No. 1.8e+02;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KXPXWR 7
 ID 72 RVPWR 78

RESULT 13
 O9F1E6 PRELIMINARY; PRT; 303 AA.

AC O9F1E6;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE SIMILARITY TO NAM.
 CS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty p1 and TAC
 RT clones."
 RL DNA Res. 7:131-135(2000).
 DR EMBL: AB026645; BAB02498.1;
 DR EMBL: AP000385; BAB02498.1; JOINED.
 DR InterPro: IPR003441; NAM.
 DR Pfam: PF02365; NAM; 1.
 SQ SEQUENCE 303 AA; 35198 MW; FDF77F62358BDECC CRC64;

Query Match 83.3%; Score 30; DB 10; Length 303;
 Best Local Similarity 57.1%; Pred. No. 2.1e+02;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KXPXWR 7
 ID 191 KYTPWR 197

RESULT 14
 O80627 PRELIMINARY; PRT; 317 AA.

AC O80627;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE PUTATIVE LYOPHOSPHOLIPASE.
 GN F1216.6.
 CS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Kounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
 RA Seearville C.R., Venter J.C.;
 RT "Arabidopsis thaliana chromosome 11 BAC F1216 genomic sequence."
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC 11- SIMILARITY TO ALPHA/BETA HYDROLASE FOLD.
 DR EMBL: AC004218; AAC27831.1;
 DR InterPro: IPR000073; Abhydrolase.
 DR InterPro: IPR000379; Est_lip_thioest_actsite.
 DR Pfam: PF00561; abhydrolase; 1.
 SQ SEQUENCE 317 AA; 35937 MW; E6429D3BD07FDCHE CRC64;

Query Match 83.3%; Score 30; DB 10; Length 317;
 Best Local Similarity 57.1%; Pred. No. 2.2e+02;
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 KFXPXWR 7
Db 167 KFIPTWK 173

RESULT 15
Q910K8 PRELIMINARY; PRT; 389 AA.
AC Q910K8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN PA2630.
GN PA2630.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae.
QC Pseudomonas.
OX NCHI_TaxID=287;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.C., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT *Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.*;
RL Nature 406:959-964(2000).
DR EMBL; AE004691; AAC06018.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 389 AA; 44057 MW; 10FFBC62E92467C7 CRC64;

```

```

Query Match      83.3%; Score 30; DH 2; Length 389;
Best Local Similarity 57.1%; Pred. No. 2.7e+02;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 KFXPXWR 7
Db 110 KFLPSWR 116

```

Search completed: February 27, 2002, 11:54:09
Job time: 229 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 27, 2002, 11:50:37 ; Search time 32.48 Seconds
(without alignments)
4.850 Million cell updates/sec

Title: US-09-446-109A-24
Perfect score: 36
Sequence: 1 KFXPXWR 7

Scoring table: HQSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	DB ID	Description
1	28	77.8	401 3	US-08-974-022-6
2	28	77.8	401 4	US-09-042-785A-12
3	28	77.8	401 4	US-08-795-445A-6
4	28	77.8	401 4	US-08-795-447A-6
5	28	77.8	401 4	US-08-974-186-6
6	28	77.8	401 4	US-08-795-446B-6
7	28	77.8	401 4	US-09-153-927-1
8	28	77.8	527 3	US-08-907-229-2
9	28	77.8	1536 4	US-09-413-814-10
10	28	77.8	2232 4	US-09-091-219-25
11	28	77.8	2247 4	US-09-091-219-2
12	27	75.0	6 1	US-08-215-137-13
13	27	75.0	456 4	US-09-268-364-21
14	27	75.0	873 3	US-08-187-331-6
15	27	75.0	925 2	US-08-382-946-1
16	27	75.0	925 2	US-08-504-169-1
17	27	75.0	925 5	PCT-US94-14893-1
18	27	75.0	1220 1	US-08-158-232-43
19	27	75.0	1220 2	US-08-611-928-43
20	27	75.0	1220 3	US-09-173-891-43
21	27	75.0	1229 3	US-09-310-293-2
22	27	75.0	1289 1	US-07-876-280-4
23	27	75.0	1289 1	US-07-675-772-4
24	27	75.0	1289 1	US-08-063-170-4
25	27	75.0	1289 1	US-08-158-232-4
26	27	75.0	1289 1	US-08-304-626-4
27	27	75.0	1289 1	US-08-316-301A-4

25 27 75.0 1289 2 US-08-611-928-4 Sequence 4, Appli
26 27 75.0 1289 3 US-09-173-891-4 Sequence 4, Appli
27 27 75.0 1289 4 US-09-076-137-4 Sequence 4, Appli
28 27 75.0 1289 5 PCT-US92-03624-4 Patent No. 5281530
29 27 75.0 1289 6 5281530-3 Patent No. 5426049
30 27 75.0 1289 6 5426049-4 Sequence 2, Appli
31 27 75.0 1385 1 US-07-876-280-2 Sequence 2, Appli
32 27 75.0 1385 1 US-07-675-772-2 Sequence 2, Appli
33 27 75.0 1385 1 US-08-063-170-2 Sequence 2, Appli
34 27 75.0 1385 1 US-08-158-232-2 Sequence 2, Appli
35 27 75.0 1385 1 US-08-304-626-2 Sequence 2, Appli
36 27 75.0 1385 1 US-08-316-301A-2 Sequence 2, Appli
37 27 75.0 1385 2 US-08-611-928-2 Sequence 2, Appli
38 27 75.0 1385 3 US-09-173-891-2 Sequence 2, Appli
39 27 75.0 1385 4 US-09-076-137-2 Sequence 2, Appli
40 27 75.0 1385 5 PCT-US92-03624-2 Patent No. 5281530
41 27 75.0 1385 6 5281530-1 Patent No. 5426049
42 27 75.0 1385 6 5426049-1

ALIGNMENTS

RESULT
US-08-174-022-6
Sequence 6, Application US/08974022
Patent No. 6015938
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,022
FILING DATE: 12-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
PRIORITY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
GS-04-474-022-6

Query Match 77.8% Score 28; DB 3; Length 401;
Best Local Similarity 66.7%; Pred. No: 2.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
CY KFXPXW 6
DB KFTPNW 219

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXW 6
|||

DB 214 KETPNW 219

RESULT 5

US-08-974-186-6
; Sequence 6, Application US/08974186
; Patent No. 6284740

GENERAL INFORMATION:

APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:

ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavenland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/974,186

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION NUMBER: 08/577,788

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 401 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-974-186-6

Query Match 77.8%; Score 28; DB 4; Length 401;

Best Local Similarity 66.7%; Pred. No. 2.1e+02;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXW 6

|||

DB 214 KETPNW 219

RESULT 6

US-08-795-446B-6

; Sequence 6, Application US/08795446B

; Patent No. 6288032

GENERAL INFORMATION:

APPLICANT: Boyle, William J.

APPLICANT: Lacey, David L.

APPLICANT: Calzone, Frank J.

APPLICANT: Chang, Ming-Shi

TITLE OF INVENTION: OSTEOPROTEGERIN

NUMBER OF SEQUENCES: 53

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amgen Inc.

STREET: 1840 Dehavenland Drive

CITY: Thousand Oaks

STATE: California

COUNTRY: USA

ZIP: 91320-1789

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/795,446B

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/577,788

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Winter, Robert B.

REFERENCE/DOCKET NUMBER: A-378

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 401 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-795-446B-6

Query Match

Best Local Similarity 77.8%; Score 28; DB 4; Length 401;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXW 6

|||

DB 214 KETPNW 219

RESULT 7

US-09-153-927-1

; Sequence 1, Application US/09153927A

; Patent No. 6297022

GENERAL INFORMATION:

APPLICANT: McDonnell, Peter C.

APPLICANT: Young, Peter R.

APPLICANT: Zol, Jun

TITLE OF INVENTION: A Method of Identifying Agonists and

TITLE OF INVENTION: Antagonists for Tumor Necrosis Related Receptors TR1, TR3

TITLE OF INVENTION: and TR5

FILE REFERENCE: GH50031

CURRENT APPLICATION NUMBER: US/09/153,927A

CURRENT FILING DATE: 1998-09-16

EARLIER APPLICATION NUMBER: 60/061,334

EARLIER FILING DATE: 1997-10-08

NUMBER OF SEQ ID NOS: 11

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 401

TYPE: PRT

ORGANISM: Human

US-09-153-927-1

Query Match

Best Local Similarity 77.8%; Score 28; DB 4; Length 401;

Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KFXPXW 6

|||

DB 214 KETPNW 219

RESULT 8

US-08-987-229-2

; Sequence 2, Application US/08907229A

```

: Patent No. 6072048
: GENERAL INFORMATION:
: APPLICANT: RILEY, Lee W.
: TITLE OF INVENTION: DNA MOLECULE ENCODING FOR CELLULAR UPTAKE OF
: MYCOBACTERIUM TUBERCULOSIS AND USES THEREOF
: FILE REFERENCE: 19603/1531
: CURRENT APPLICATION NUMBER: US/08/907,229A
: CURRENT FILING DATE: 1997-08-06
: EARLIER APPLICATION NUMBER: 60/040,097
: EARLIER FILING DATE: 1997-03-10
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 2
: LENGTH: 527
: TYPE: PRT
: ORGANISM: Mycobacterium tuberculosis
: US-08-907-229-2

```

```

Query Match      77.8%; Score 28; DB 3; Length 527;
Best Local Similarity 66.7%; Pred. No. 2.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 FXPXWR 7
      | | | |
DB      3 FGPSWR 8

```

```

RESULT 9
US-09-413-814-10
: Sequence 10, Application US/09413814
: Patent No. 6225064
: GENERAL INFORMATION:
: APPLICANT: Gesellschaft fuer Hiotechnologische Forschung mbH
: APPLICANT: Bristol-Myers Squibb, Co.
: APPLICANT: Beyer, Stefan
: APPLICANT: Bioecker, Helmut
: APPLICANT: Brandt, Petra
: APPLICANT: Cino, Paul M
: APPLICANT: Dougherty, Brian A
: APPLICANT: Goldberg, Steven L
: APPLICANT: Hofle, Gerhard
: APPLICANT: Mueller, Joachim
: APPLICANT: Reichenbach, Hans
: TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
: heteropolypeptide compounds
: FILE REFERENCE: PCT/US 99/23535
: CURRENT APPLICATION NUMBER: US/09/413,814
: CURRENT FILING DATE: 1999-10-07
: EARLIER APPLICATION NUMBER: DE 198 46 493.2
: EARLIER FILING DATE: 1998-10-09
: NUMBER OF SEQ ID NOS: 107
: SOFTWARE: Patentin Ver. 2.1
: SEQ ID NO 10
: LENGTH: 1536
: TYPE: PRT
: ORGANISM: Sorangium cellulosum
: US-09-413-814-10

```

```

Query Match      77.8%; Score 28; DB 4; Length 1536;
Best Local Similarity 66.7%; Pred. No. 7.7e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 FXPXWR 7
      | | | |
DB     896 FLPAWR 901

```

```

RESULT 10
US-09-091-219-25
: Sequence 25, Application US/09091219
: Patent No. 6171592

```

```

: GENERAL INFORMATION:
: APPLICANT: STUDDERT, Michael J.
: APPLICANT: CRABB, Brendan S.
: APPLICANT: FENG, Li
: TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
: FILE REFERENCE: 040268/0151
: CURRENT APPLICATION NUMBER: US/09/091,219
: CURRENT FILING DATE: 1998-10-05
: EARLIER APPLICATION NUMBER: PCT/AU96/00815
: EARLIER FILING DATE: 1996-12-18
: EARLIER APPLICATION NUMBER: AU PN7201
: EARLIER FILING DATE: 1995-12-18
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 25
: LENGTH: 2232
: TYPE: PRT
: ORGANISM: equine rhinovirus 1
: US-09-091-219-25

```

```

Query Match      77.8%; Score 28; DB 4; Length 2232;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 FXPXWR 7
      | | | |
DB     2227 FVPTWR 2232

```

```

RESULT 11
US-09-091-219-2
: Sequence 2, Application US/09091219
: Patent No. 6171592
: GENERAL INFORMATION:
: APPLICANT: STUDDERT, Michael J.
: APPLICANT: CRABB, Brendan S.
: APPLICANT: FENG, Li
: TITLE OF INVENTION: EQUINE RHINOVIRUS 1 PROTEINS
: FILE REFERENCE: 040268/0151
: CURRENT APPLICATION NUMBER: US/09/091,219
: CURRENT FILING DATE: 1998-10-05
: EARLIER APPLICATION NUMBER: PCT/AU96/00815
: EARLIER FILING DATE: 1996-12-18
: EARLIER APPLICATION NUMBER: AU PN7201
: EARLIER FILING DATE: 1995-12-18
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 2
: LENGTH: 2247
: TYPE: PRT
: ORGANISM: equine rhinovirus 1
: US-09-091-219-2

```

```

Query Match      77.8%; Score 28; DB 4; Length 2247;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      2 FXPXWR 7
      | | | |
DB     2227 FVPTWR 2232

```

```

RESULT 12
US-08-215-137-13
: Sequence 13, Application US/08215137
: Patent No. 5614370
: GENERAL INFORMATION:
: APPLICANT: Konteatis, Zenon
: APPLICANT: Siciliano, Salvatore J
: APPLICANT: Springer, Martin S
: TITLE OF INVENTION: ASSAY TO IDENTIFY HUMAN C5A ANTAGONISTS

```

us-09-446-109a-24.ra1

Thu Feb 28 11:24:15 2002

TITLE OF INVENTION: PLANT GLYCOLYSIS AND RESPIRATION ENZYMES
REFERENCE: BB-1154
CURRENT APPLICATION NUMBER: US/09/268,364A
CURRENT FILING DATE: 1999-03-15
EARLIER APPLICATION NUMBER: 60/079,387
EARLIER FILING DATE: March 16, 1998
NUMBER OF SEQ ID NOS: 21
SOFTWARE: Microsoft Office 97
SEQ ID NO 21
LENGTH: 456
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-251-364-21

Query Match 75.0%; Score 27; DB 4; Length 456;
Best Local Similarity 66.7%; Pred. No. 3.8e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 FXPXR 7
DB 209 EGPWR 214

RESULT 14
US-09-267-331-6
Sequence 6, Application US/09187331
Patent No. 6043056
GENERAL INFORMATION:
APPLICANT: Yoo, Henry
APPLICANT: Corley, Neil C.
APPLICANT: Guillet, Karl J.
APPLICANT: Gorgone, Gina A.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: CELL SURFACE GLYCOPROTEINS
FILE REFERENCE: PF-0631 US
CURRENT APPLICATION NUMBER: US/09/187,331
CURRENT FILING DATE: 1998-11-06
NUMBER OF SEQ ID NOS: 6
SOFTWARE: PERL Program
SEQ ID NO 6
LENGTH: 873
TYPE: PRT
ORGANISM: Homo sapiens
SEQUENCE:
CHECK INFORMATION: g189650
US-09-267-331-6

Query Match 75.0%; Score 27; DB 3; Length 873;
Best Local Similarity 66.7%; Pred. No. 7.1e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY KFXPXW 5
DB 1111
KFXPXW 248

RESULT 15
US-09-268-364-1
Sequence 1, Application US/08392946
Patent No. 5939269
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: The Regents of the University of California
APPLICANT: Goldfine, Ira D.
APPLICANT: Grope, Andrew
APPLICANT: Maddux, Betty A.
APPLICANT: Shencer, Steven
APPLICANT: Stewart, Timothy A.
TITLE OF INVENTION: Antagonists to Insulin Receptor Tyrosine
KINASE
FILE REFERENCE: Kinase Inhibitor
CURRENT APPLICATION NUMBER: 1

TITLE OF INVENTION: AND ACONISTS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merck & Co., Inc.
STREET: 126 Lincoln Ave., P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: United States of America
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/215,137
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Hencen, Gerard H
REGISTRATION NUMBER: 35,746
REFERENCE/DOCKET NUMBER: 19108
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3901
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 6 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: C-terminal
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1
OTHER INFORMATION: /label= BII
OTHER INFORMATION: /note= "either the natural phenylalanine amino
terminus or the Bolton-Hunter modified peptide
having the group 3-(p-hydroxyphenyl)propionyl group
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /label= dCha
OTHER INFORMATION: /note= "D-cyclohexylalanine"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 6
OTHER INFORMATION: /label= dArg
OTHER INFORMATION: /note= "D-arginine"
US-08-215-137-13

Query Match 75.0%; Score 27; DB 1; Length 6;
Best Local Similarity 83.3%; Pred. No. 1.6e+05;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 FXPXR 7
DB 1111
FXPXW 6

RESULT 13
US-09-268-364-21
Sequence 21, Application US/09268364A
Patent No. 6204063
GENERAL INFORMATION:
APPLICANT: Allen, Steve
APPLICANT: Lee, Jian Ming
APPLICANT: Lightner, Jonathan
APPLICANT: Oucil, Joan

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genentech, Inc.
: STREET: 1 DNA Way
: CITY: South San Francisco
: STATE: California
: COUNTRY: USA
: ZIP: 94080
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WinPalin (Genentech)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/392,946
: FILING DATE:
:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/182241
: FILING DATE: 14-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Kubinec, Jeffrey S.
: REGISTRATION NUMBER: 36,575
: REFERENCE/DOCKET NUMBER: P0875P1PC1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650/225-8228
: TELEFAX: 650/952-9881
: INFORMATION FOR SEQ ID NO: 1:
: LENGTH: 925 amino acids
: TYPE: Amino Acid
: TOPOLOGY: Linear
:
: 'S-08-392-946-1

```

```

Query Match      75.0%; Score 27; DB 2; Length 925;
Best Local Similarity 66.7%; Pred. No. 7.5e+02;
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFXPW 6
DB 295 KFNPM 300

```

```

Search completed: February 27, 2002, 11:50:38
Job time: 867 sec

```